How symbolic and embodied representations work in concert

Hutchinson, Sterling

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

How symbolic and embodied representations work in concert

Sterling Hutchinson

How symbolic and embodied representations work in concert

Sterling Hutchinson
Ph.D. thesis
Tilburg University

TiCC Ph.D. series no. 39

ISBN: 978-94-6203-862-2
Print: CPI-Wöhrmann Print Service – Zutphen
Cover design: Wesley Klehm

How symbolic and embodied representations work in concert

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan Tilburg University,
op gezag van de rector magnificus,
prof. dr. E. H. L. Aarts,
in het openbaar te verdedigen ten overstaan van een
door het college voor promoties aangewezen commissie
in de aula van de Universiteit
op dinsdag 30 juni 2015 om 10:15 uur

door

Sterling Chelsea Hutchinson

geboren op 10 maart 1988 te Altamonte Springs, FL, USA

Promotores:
Prof. Dr. M. M. Louwerse
Prof. Dr. E. O. Postma

Promotiecommissie:
Dr. L. Connell
Dr. H. IJzerman
Prof. Dr. A. van den Bosch

Chapter One
Introduction

Introduction

Think about the first time you read *Peter Pan* as a child. The words probably came to life as you imagined Peter and his friends flying past a huge wooden pirate ship, and you could almost hear the ticking from the crocodile in the bay, just as if the story was part of the real world. At the same time, it is unlikely you considered how the series of words on a page came together into a rich and meaningful story. This very question regarding the nature of cognition, and specifically that of language comprehension and how we understand words on a page has been debated for years. Researchers have argued that the format of mental representations is either inherently linguistic (i.e., amodal symbols; Fodor, 1975) or inherently perceptual (i.e., modal embodied states; Barsalou, 1999; Glenberg, 1997) in nature. This debate has recently been infused with a new perspective that examines mental representations from a combined point of view (Barsalou, 2008). This new perspective endorses neither linguistic nor perceptual accounts of representation but rather suggests that both types of representations work together to facilitate language comprehension. Instead of spending time and resources exploring *whether* mental representations are linguistic or perceptual it is now more productive to explore the question of *when* mental representations are more linguistic or more perceptual. This distinction is significant, as it asks an informative question that leads to exciting new research questions that contribute to a range of disciplines in the cognitive sciences, including linguistics, psychology, and computer science. This dissertation focuses on some of those questions by asking how language processing is facilitated by both symbolic and embodied accounts working in concert.

Embodied Theories of Cognition

Proponents of embodied or perceptual representations have offered a modal view of cognition, stating that comprehension is driven by perceptual experiences so much so that words and concepts are grounded in the physical world through action and perception (Barsalou, 1999; Glenberg, 1997; Zwaan, 2004). In essence, the mind is embodied and word meaning must be grounded in bodily experiences and situations. The activation of word and concept meaning in memory is derived from modality-specific re-enactments or simulations of the external experiences associated with those concepts. For example, when you read the word *thimble* on a page, it is understood by simulating the same patterns of neural activation that are active when seeing the size, shape, material, and color of a thimble, touching the hard and pitted metal surface of a thimble, and using granny's favorite thimble to protect your finger while sewing. All of these perceptual experiences come together to represent the many facets of a thimble and contribute to word meaning, allowing us to understand just what exactly Wendy handed to Peter in lieu of a kiss. According to this embodied perspective, mental representations are simply mental reenactments or simulations of external percepts.

Theorists have hypothesized that embodied representations are fundamental to language processing (Barsalou, 1999; Glenberg, 1997; Pecher & Zwaan, 2005; Semin & Smith, 2008). Indeed, these claims are backed by ample empirical support. For example, language processing is facilitated when experimental tasks allude to perceptual features related to stimuli. These perceptual features are numerous, with location, perspective, orientation, shape, color, direction, and even the modality of stimuli impacting language processing.

According to this modal view of cognition, simulated location is critical during language comprehension. For example, spatially relevant word pairs presented in their expected physical locations are processed faster than when they are presented in an unexpected location. This occurs because a word is easier to process when the anticipated and actual perceptual properties of the word match. Zwaan and Yaxley (2003) demonstrated this in an experiment in which they showed word pairs such as *attic* and *basement* in a vertical configuration. Response times (RTs) were faster when *attic* appeared above *basement* than vice versa. In similar studies, Šetić and Domijan (2007) and Pecher, van Dantzig, Boot, Zanolie, and Huber (2010) presented 'up' and 'down' words (e.g., a flying animal like an *eagle* or an animal that cannot fly like a *dolphin*) one at a time either in an expected physical location or in an unexpected physical location. Participants were faster to process concept and location matches than mismatches. Likewise, Estes, Verges, and Barsalou

(2008) found the same effects with an object word (e.g., *cowboy*) followed by a high or low location cue (e.g., *hat* versus *boot*). As predicted, participants were faster to identify a target letter appearing in a location that matched the cue. Bergen, Lindsay, Matlock, and Narayanan (2007) found similar effects when participants listened to sentences implying 'up' or 'down' motions. When sentences matched the position of a visual shape cue, processing was facilitated. Richardson, Spivey, Barsalou, and McRae (2003) found that even abstract verbs like *argue* and *respect* are related to specific orientations (horizontal for *argue* and vertical for *respect*) and expectedly RTs are influenced accordingly when visual stimuli are oriented horizontally or vertically.

The same is true for perspective, Borghi, Glenberg and Kaschak (2004) showed that when participants read sentences implying a perspective (e.g., "You are eating in a restaurant" or "You are waiting outside a restaurant") RTs to a concept verification task were faster if the perspective of the concept and sentence matched (e.g., *table* would be a match with "You are eating in a restaurant" whereas it would be a mismatch with "You are waiting outside a restaurant"). This effect is also explained in terms of embodied representations whereby perceptual simulations of the words presented are automatically generated and relied upon by participants during the experimental task.

Embodied effects extend past linguistic stimuli to also include pictorial stimuli, with pictures that appear in their expected spatial positions similarly being processed faster than mismatched picture-location pairs (Crawford, Margolies, Drake, & Murphy, 2006; Dijkstra, Yaxley, Madden, & Zwaan, 2004; Meier et al., 2007). For instance, RTs decrease when the orientation of an image matches implied physical characteristics within a related sentence. Stanfield and Zwaan (2001) demonstrated this when they presented participants with an image of an item followed by a sentence describing the item in an orientation consistent or inconsistent with the previously presented image. Participants read “John put the pencil in the cup,” versus “John put the pencil in the drawer,” and then made an item recognition judgment after seeing a picture of a vertically oriented pencil. The results showed that participants exhibited faster RTs when reading sentences describing objects in the same orientations as the pictures depicting those objects.

Zwaan, Stanfield, and Yaxley (2002) used a similar paradigm where participants read about objects in scenarios that implied a particular object shape, and RTs were faster when a presented picture matched the implied shape in the sentence. Not only can the same effect also be found for item shape but even color is perceptually simulated. Connell and Lynott (2009) showed that when a presented color word matched a particular color implied by the previous sentence, color naming was easier.

Myung, Blumstein, and Sedivy (2006) found that when a words that implied particular manipulation features were presented auditorily, participants made faster decisions about the another prime that shared manipulation features (e.g., *typewriter* and *piano*) than an unrelated prime (e.g., *blanket*). Glenberg and Kaschak (2002), Borreggine and Kaschak (2006), and Bergen and Wheeler (2005), comparably showed that sentences that implied directional movements were processed faster when the response required a congruent motion instead of an incongruent motion. For instance, when participants read, “Open the drawer” and were asked to respond by making a motion towards their body, RTs were faster than when reading a sentence like “Close the drawer.” In essence, understanding “Open the drawer” is thought to generate simulations of actions toward the body, thereby facilitating a response motion in the same direction. Similarly, Zwaan and Taylor (2006) also asked participants to read sentences implying motion, such as “He turned down the volume," and they found that when response actions, such as rotating a knob to the left versus rotating a knob to the right, were congruent with what was described in the sentence, responses were facilitated.

This paradigm has been replicated and extended to less concrete words, with researchers finding similar effects for verbs (Meteyard, Zokaei, Bahrami, & Vigliocco, 2008), positive/negative metaphors (Meier & Robinson, 2004), abstract concepts (Meier, Hauser, Robinson, Friesen, & Schjeldahl, 2007), and powerful/powerless metaphors (Schubert, 2005) to name just a few. For example, Meteyard, Zokaei, Bahrami, and Vigliocco (2008) found that motion verbs (e.g., *rise*, *fall*) were processed faster when the motion implied from the word matched the direction of a simultaneously presented visual motion pattern. On the other hand, Kaschak, Madden, Therriault, Yaxley, Aveyard, Blanchard, and Zwaan (2005) found a mismatch advantage when participants were processing sentences implying motion (e.g., “The dog was running towards you” or “You backed away from the fire”) while simultaneously perceiving motion in the opposite direction, presumably because the neural network required for simulation was occupied with a perceptual task.

Even metaphorically high and low words such as *God* and *devil*, *good* and *bad,* or *strong* and *weak,* were processed faster when presented in a congruent vertical position on the screen (Meier, Hauser, Robinson, Friesen, & Schjeldahl, 2007; Meier & Robinson, 2004; Schubert, 2005). Santana and de Vega (2011) found that conceptual metaphors are processed through embodied mechanisms and Wilson and Gibbs (2007) found that action specifically impacted metaphor comprehension. They found when asked to perform particular actions, participants were faster to comprehend metaphors when the metaphor-action pair was matched. Findings like these provide overwhelming evidence for theories that endorse embodied representations in both literal as well as figurative language (Glenberg & Kaschak, 2002; Lakoff, 1987; van Dantzig, Pecher, Zeelenberg & Barsalou, 2008).

Such behavioral RT evidence for embodied cognition dovetails with neuropsychological studies that show sensory and motor activation during language processing (Buccino et al., 2005; Kan et al., 2003; Rueschemeyer et al., 2010). Researchers have identified neural activation in the motor cortex when individuals read words conveying relevant motor-driven actions.

Participants read sentences that included action words related to the legs (e.g., *kick*), the face (e.g., *lick*), or the arms (e.g., *pick*). When those action words were read, neural activity in the corresponding motor cortex area for actual movement of those body parts increased, suggesting that action verbs are understood through embodied mechanisms (Hauk, Johnsrude, & Pulvermuller, 2004). In a transcranial magnetic stimulation (TMS) study, Buccino et al. (2005) showed that when participants listen to hand-related or foot-related action sentences (e.g., “he sewed the skirt” versus “he kicked the ball”) the relevant portion of the motor system was activated. Chao and Martin (2000) even found that motor cortex activity increased significantly when participants were presented with pictures of highly manipulable objects compared to un-manipulable objects.

Based on this evidence, embodied representations are thought to be modality specific with visual information activating visual cortex, tactile information activating motor/somatosensory cortex, and emotion information activating emotional centers, leading to faster processing times within the same modality than across different modalities (e.g., Marques, 2006; Spence, Nicholls, & Driver, 2001). Indeed, processing within specific modalities is faster than processing between modalities (Van Dantzig, Pecher, Zeelenberg, & Barsalou, 2008). For instance, Pecher, Zeelenberg, and Barsalou (2003) found that when verifying facts (e.g., leaves can rustle, or cranberries are tart) switching between perceptual modalities took longer than when both facts involved the same sensory modality. These findings suggest when we mentally simulate a sentence, we are activating a neural pattern of activation that is modality specific.

Studies like these (see Barsalou, 2008; Pecher & Zwaan, 2005; Semin & Smith, 2008 for more complete overviews) demonstrate that individuals rely on perceptual representations in everyday language comprehension. The results from these experiments, as well as other studies like them (Glenberg, & Kaschak, 2002; Pecher, van Dantzig, Zwaan, & Zeelenberg, 2009; Spivey & Geng, 2001; Zwaan & Yaxley, 2003), show that participants' processing appears to benefit from experimental tasks and activities that invoke the consideration of perceptual information. Such findings support the idea that embodiment and perceptual simulation is central to cognition and provide evidence that language processing is facilitated by the use of embodied mental representations.

Symbolic Theories of Cognition

The embodied cognition models of human thought gained popularity in response to classical amodal symbolic cognition accounts that dominated in the 1950s — 1970s. Symbols are non-iconic representations of their referents. Although there are several interpretations of a symbolic cognition account, the typical account proposed in contrast to grounded cognition suggests that we make amodal symbolic connections between a concept and its meaning in memory (Fodor, 1975; Pylyshyn, 1984; Tulving, 1983). Meaning can then be derived from the symbolic linguistic connections that exist between symbols in a network (Tulving & Thomson, 1973). In other words, concepts are represented in our minds in a propositional way through amodal lists and semantic networks, and word meaning is established from relationships between symbols without requiring any perceptual simulations to garner meaning (Kintsch, 1998; Landauer & Dumais, 1997; Pylyshyn, 1984).

In essence, words need not activate modality specific sensorimotor simulations in order to be understood. For example, computer systems are amodal; keystrokes and mouse movements are translated into ones and zeroes that do not directly represent the commands being provided. In a simplified conceptualization of this account, the word *thimble* can be thought of as being translated into 01110100 01101000 01101001 01101101 01100010 01101100 01100101 by our mind. This abstract symbolic representation is a sort of mental translation from a concept in the external world to a mental representation. Unlike an embodied account, *thimble* is not directly related to an actual thimble, or a perceptual re-enactment of a thimble. It isn't necessary to imagine what a thimble feels or looks like, symbolically, humans and computers alike can determine the meaning of a given word, e.g., *thimble*, through assessing the strength of statistical relationships between that word (i.e., *thimble*) and what is related to that word (e.g., *sewing*, *finger*, *metal*, *cap*, *thread, needle, protection)*.

Symbolic systems function by calculating covariation between words, features, and concepts from which meaning emerges. Because strictly symbolic representations do not correspond to perceptual states they instead rely on mathematical and computational algorithms for symbol manipulation to generate meaning. This makes it easy to represent abstract concepts in a symbol-type system. In fact, it is easy to imagine how situations that are impossible to ever experience can be represented symbolically (Pylyshyn, 2002). Consider again Peter Pan; while it is not possible to stop aging, and especially not possible for a human to fly, we have little trouble imagining either of these scenarios. Scenarios like these, and abstract concepts (e.g., *infinity*) are not based on previous perceptual or bodily experiences but they can easily be represented as a node or symbol in a network of connections. In line with this example, language comprehension theorists suggest that this is how text is understood. Mental representations are thought of as arbitrary and abstract symbols, with sentences being understood as a network of related but amodal propositional units (Kintsch, 1998; Van Dijk & Kintsch, 1983). As such, symbolic representations lend themselves to computational and statistical processing to represent knowledge, a process that is more efficient than taking the time to mentally re-enact perceptual experiences, as modal experience is all more or less encoded in the same abstract fashion.

Working computer programs and computational language models and algorithms are often used as examples of symbolic systems (e.g., ACT; Anderson, 1996; or LSA; Landauer & Dumais, 1997). For instance, Latent Semantic Analysis (LSA) uses large corpora to compute word meaning by mapping words and their neighbors in a high dimensional semantic space. Words or texts in this semantic space are then compared in terms of whether they appear in the same and similar contexts and the similarity between them is calculated and represented by a cosine value. LSA has shown to approximate human performance in a number of ways. Despite LSA being unable to understand what words mean in the same way humans seem to, LSA is still quite successfully able to algorithmically compute word meaning. For example, LSA has shown, on the basis of symbolic algorithms alone, to be able to pass the Test of English as a foreign language (TOEFL) test (Landauer & Dumais, 1997). LSA has also been able to grade essays just as well as expert graders (Landauer, Foltz, & Laham, 1998).

Although some of the best examples of symbolic systems seem to be computational models, evidence suggests that humans also use statistical regularities in language to establish word meaning. Symbolic representations are often framed as being contrary to embodied cognition research (de Vega et al., 2008; Glenberg, 2010; Lakens, 2011; Louwerse, 2011a). Yet, there is evidence that symbolic representations (i.e., language statistics) actually encode perceptual information about the world around us (Louwerse, 2008). Experiments and computer simulations demonstrate that individuals rely on symbolic linguistic representations in everyday language comprehension. For instance, pairs of words, such as *up–down or top–bottom* occur more frequently when a concept in a high location precedes a concept word in a low location

(Louwerse, 2008) and the same holds for a variety of paired words, with word frequency patterns matching their perceptual relations in the real world. For example, concepts like body parts are processed in the same way. Co-occurrence frequencies of name pairs can predict the vertical position of body parts, so *head* appears before *shoulder* in language, just as heads appears above the shoulders in real life (Tillman, Hutchinson, & Louwerse, 2013).

Statistical regularities are not only used when processing perceptually related words. Hutchinson and Louwerse (2012) showed that language statistics also explain metaphor processing, with positive words (e.g., *achievement*, *beautiful*) appearing before negative words (e.g., *failure*, *ugly*). Further, the linguistic frequencies of how often the word pairs occur in each order (i.e., *beautiful* appearing before *ugly*, or *ugly* appearing before *beautiful*) predicted participant RTs in a semantic judgment task. This pattern for metaphor word pairs was further extended to concepts related to temperature, authority, and gender (Hutchinson & Louwerse, 2013). Tillman, Hutchinson, Jordan, and Louwerse (2013) extended this effect to emotional information as well, with linguistic frequencies of emotional nouns and adjectives (e.g., *birthdays* can be *happy*, *insults* can be *devastating*) predicting whether the two words shared the same emotional valence. Therefore even though pairs like *eagle* — *dolphin* are processed faster than the reverse, it might just be the case that we are processing these words symbolically.

Louwerse, Cai, Hu, Ventura, and Jeuniaux (2006) and Louwerse and Zwaan (2009) provide further evidence that humans rely on word frequencies by showing that language encodes geographical information. Louwerse and Zwaan (2009) showed that statistical linguistic frequencies between cities in the USA correlate with the actual physical distance between them. Even more, the geographical location of cities can be predicted based on whether city names tended to appear in similar linguistic contexts. Louwerse and Zwaan (2009) also showed that the latitude and longitude of the 50 largest cities in the USA could be calculated by their co-occurrence frequencies in the English language. Louwerse, Hutchinson, and Cai (2012) found the same for cities in China and the Middle East and Davies (2013) extended the finding to the UK. Louwerse and Benesh (2012) even demonstrated that fictional city locations in the *Lord of the Rings* trilogy could be predicted based on the computational semantic associations between cities in the text. Tillman, Hutchinson, and Louwerse (2013) also showed that language users rely on statistical regularities when considering geographical information, with northern and western city names appearing above/before southern and eastern city names not only in the real world, but also in language.

Just like geographical information, Hutchinson, Datla, and Louwerse (2012) show that social information is also inherent computationally in language. Social proximity and relationships between characters found in the *Harry Potter* novels can be computed via statistical algorithms. In fact, these computations can approximate human performance when generating a social network. These findings that language users rely on statistical information are not conclusive, but at least they do indicate that symbolic representations play a role in language processing. In sum, participants' processing benefits not only from the consideration of perceptual embodied features as previously presented but also from statistical regularities in language.

Integrated Theories of Cognition

The aforementioned evidence traditionally suggests two contrary conclusions: a) that language processing is embodied or b) that language processing is symbolic. Each approach is informative because each highlights the roles that embodied and statistical factors play during language processing, but these accounts are often presented as mutually exclusive (see de Vega, Glenberg, & Graesser, 2008 for an overview of this debate). It is clear many researchers consider these two explanations to be dichotomous, with hundreds of journal articles devoted to the question of whether or not perceptual simulations plays a role in language comprehension. In the literature, this sentiment is made explicit, with researchers stating that "[symbolic] notations […] constitute a problem for the question how symbols are given meaning" (van Dantzig, Pecher, Zeelenberg & Barsalou, 2008, p. 580).

Despite the fact that symbolic and embodied accounts of cognition are so often framed as being divergent from one another (De Vega, Glenberg &

Robertson, 2008; Fodor, 2008; Glenberg, 2010; Glenberg & Robertson, 2000; Van Dantzig et al. 2008), there need not be such a division. A unified account offers resolutions for discrepancies in each account while still being mutually reinforcing (Andrews, Vigliocco, & Vinson, 2009; Dove, 2009). In fact there is increasing evidence that both linguistic processes and simulation processes both play a role during language processing, and that symbolic and embodied cognition accounts can be integrated (Louwerse, 2008; Louwerse, 2011b). That is, statistical linguistic factors and perceptual simulations interact with one another such that linguistic representations are used as external symbols to facilitate processing. Several researchers have already proposed that it is important to consider the interplay between symbolic and embodied factors in cognition (Barsalou, Santos, Simmons, & Wilson, 2008; Louwerse, 2008, 2010; Louwerse & Jeuniaux, 2008, 2010; Zwaan 2014).

One such theory is Paivio's (1971; 1986) Dual Coding Theory where cognitive processes include both visual and verbal information. In this theory, pictorial stimuli allow for pictorial representations, and verbal stimuli allow for linguistic representations. Paivio proposed three levels of meaning. The first level is representational, where verbal stimuli are represented as words and pictorial stimuli are represented as images. The second level is referential whereby linguistic and perceptual representations refer to one another and form connections. The third level is associative, involving intraverbal and interimaginal representations. In essence, each representation is processed along distinct channels, but relationships exist between channels such that different types of representations might be employed in different situations. For instance, this theory implies that verbal stimuli are first are foremost represented linguistically whereas pictorial stimuli are first and foremost represented perceptually. This theory also implies that neither explanation (embodied or symbolic) should be dismissed but instead both amodal linguistic information and modal perceptual information refers to one another to work together to represent meaning.

Like Paivio, Barsalou, Santos, Simmons & Wilson's (2008) Language and Situated Simulation (LASS) theory also suggest that representations are not solely perceptual. According to LASS, there are also both linguistic and simulation systems. During processing, both systems are engaged immediately, but linguistic activation is more important immediately and embodied simulation becomes important later. In a nutshell, perceptual symbols can function symbolically, being used as modal representations during linguistic computations; upon seeing a word, both linguistic and perceptual representational systems become immediately active but linguistic representations are more important immediately whereas the more relevant perceptual simulations become more important later in processing.

A similar theory, the Symbol Interdependency Theory (Louwerse 2007) proposes that mental representations are linguistic, through statistical relationships between words within language, and embodied, through the references language makes to external perceptions (Louwerse, 2007, 2008; Louwerse & Jeuniaux, 2008, 2010). According to this theory, language encodes perceptual information and we use language as a shortcut during processing. Like the LASS theory, linguistic information is important for shallow and quick mental representations, but perceptual simulations are more relevant for deeper mental representations (Louwerse & Jeuniaux, 2010). So even with limited grounding, meaning is garnered through language statistics alone. To summarize, linguistic forms depend on one another while still referring to perceptual representations, such that language encodes perceptual information. The Symbol Interdependency Theory also suggests that not every word has to be grounded in perceptual experience, as words can make reference to other related words to establish meaning.

Evidence has started to accumulate in favor of theories that consider both symbolic and embodied factors in cognition. For example, Louwerse (2008) and Tse, Kurby, and Du (2010) found that RTs to semantic judgments for words like *attic–basement* can be explained by both language statistics and perceptual simulation. Louwerse (2008) calculated statistical linguistic frequencies for each set of word pairs as well as participants' iconicity ratings of each word pair. Both of these factors explained participant RTs during a semantic judgment task, however, the statistical linguistic factor explained more variance. Results like these suggest that both symbolic (linguistic frequencies) and perceptual (iconicity ratings) play a role in language processing.

Furthering this finding, in four experiments Louwerse and Jeuniaux (2010) found that for shallow tasks, like semantic judgment tasks, linguistic relationships between word pairs explain RTs better than an iconic factor. Cognitive processing in the iconicity judgement task were deeper than in the semantic judgement task, because a prerequisite for the iconicity judgement was a semantic judgement. In the deeper iconicity judgment task, the perceptual relationship between stimuli (an iconic factor) better explained RTs. Not only was the task relevant but processing was also modified based on the type of stimuli presented, with linguistic frequencies better explaining RTs to words and iconic ratings better explaining RTs to pictures. It is important to note that in all experiments, both linguistic and perceptual factors played a role, simply the relative importance of each factor varied due to task and stimuli conditions. Put simply, for shallow mental representations, linguistic factors were more important than embodiment factors, but for deeper mental representations this was reversed (Louwerse & Jeuniaux, 2010).

Louwerse and Connell (2011) also demonstrated that symbolic representations better explained very early fast RTs, meaning that linguistic relationships provide a good enough incomplete mental representation. On the other hand, perceptual factors explained slower RTs, suggesting that a full and complete mental reenactment is most likely generated. These results indicate that participants garner just enough information from word co-occurrences to understand a general sense of word meaning but perceptual simulates are necessary to fill in the rest of the picture. Louwerse and Hutchinson (2012) confirmed these findings, also showing that that fast RTs were best explained by linguistic information, and slow RTs were best explained by perceptual information. Furthermore, they found that EEG results also showed linguistic cortical areas to be more active during a semantic task. Similarly, perceptual cortical areas were relatively more active during an iconic task where participants relied more on perceptual information. Regardless of task, neural activation began in language processing cortical areas relatively more than perceptual processing areas and later dispersed towards perceptual processing areas relatively more than language processing areas. These findings together indicate that both linguistic and perceptual representations are important parts of language processing, but that their relative importance is impacted by the constraints of the task at hand.

Such findings show that the prominence of less-precise linguistic processes (i.e., symbolic representations) precedes the prominence of more precise simulation processes (i.e., embodied representations).

So far, the question was discussed how symbolic and embodied accounts of cognition explain how linguistic symbols attain meaning. To strive towards resolving the question of the nature of mental representations we must move past presenting perceptual and symbolic accounts as mutually exclusive explanations. The specific focus of this manuscript investigates how language processing is facilitated by both symbolic and embodied accounts working in concert.

This dissertation tests how different modulators affect the activation of linguistic and embodied representations. Instead of asking *if* processing relies upon symbolic or embodied representations, the question is posed *when* linguistic and perceptual representations are more or less relevant during language processing, and under what conditions it is likely that participants will rely more on one type of representation than another. This question is explored by investigating how symbolic and embodied cognitive processes are modulated by different factors. More specifically, the question will be addressed to what extent linguistic and perceptual representations are impacted by 1) the time course of processing 2) the spatial presentation of stimuli 3) individual differences or 4) the orientation of stimuli.

Chapter 2 demonstrates that experimental results can be explained by both linguistic and embodied factors. In three experiments, I replicate the spatial–numerical association of response codes (SNARC) effect whereby responses made with participants' left hands yield faster response times (RTs) for smaller numbers than for larger numbers. This effect is traditionally explained in terms of embodied cognition with participants perceptually simulating number magnitude on a mental number line. In essence, when processing numbers, participants mentally *see* the numbers arranged from small to large. However, this is not the only explanation; I also show that the SNARC effect can be explained by language statistics, with the linguistic frequencies numbers mirroring the SNARC effect. These results demonstrate that those effects explained solely in terms of perceptual simulation can also be explained by language statistics.

In Chapter 3 this finding was extended by exploring the factors of time and space in two experiments. Then, in a third experiment I demonstrate symbolic and embodied processes are indeed independent. Experiment 1 investigated how the use of linguistic and perceptual representations was impacted when the time course of an experimental trial was constrained. Under time constraints, linguistic frequencies best accounted for participant RTs, but both linguistic and perceptual explanations account for slower RTs. Experiment 2 explores how the spatial presentation of stimuli on the screen might also impact how and when participants are more or less likely to rely on linguistic versus perceptual representations. In a RT experiment participants view physical-location words at various locations on the screen. For words presented at the top or bottom of the screen, word meaning influences RTs. But for words appearing in the center of the screen, word frequency plays a more important role. In other words, judgments about words are made relative to other words on the screen and not relative to their absolute location on the screen. Relying on the finding that linguistic processing is more important in the early stages of language processing and perceptual processing is more important later, in a third Experiment I demonstrate that both linguistic and perceptual representations, although intertwined, are relied upon to differing extents based on the nature of the relationship shared between word pairs. In a single RT experiment where participants determine whether linguistically and/or perceptually similar or dissimilar word pairs are semantically related, linguistically related pairs are processed faster than pairs lacking a linguistic relationship whereas perceptually related and unrelated word pairs take longer to process, implying perceptual representation. Furthermore, word frequency predicts RTs for semantically related pairs, whereas both word frequency and perceptual factors are necessary to predict performance for perceptually related pairs. Importantly, for unrelated word pairs, perceptual factors alone predict RT performance, suggesting that a full perceptual representation is independently utilized when generating a relationship for unrelated word pairs. Together these findings show that language processing is both linguistic and embodied, but to different extents in different situations.

In Chapter 4, I discuss how the degree to which linguistic and perceptual information contribute to mental representations varies based on the orientation of the stimuli and on individual differences. Even though previous research has argued that primary metaphor processing can best be explained by an embodied cognition account, in four experiments I show that language statistics can explain the processing of primary metaphors that share an embodied vertical relationship (e.g., *X* above *Y* or *Y* above *X*). Furthermore, these linguistic effects were modified by participant gender, with female participants being more sensitive to statistical linguistic context than male participants.

Chapter 5 examines effect sizes computed from 126 experiments in 51 previously published embodied cognition studies to clarify the conditions under which perceptual simulations are most important. That effects of language statistics tend to be as large or larger than those of perceptual stimulation and factors associated with immediate processing (button press, word processing) reduced the effect size of perceptual simulation. These findings are considered in respect to the Symbol Interdependency Hypothesis, which argues that language encodes perceptual information, with language statistics explaining quick, good-enough representations and perceptual simulation explaining more effortful, detailed representations.

Finally, in Chapter 6 I present a brief discussion where I present several mathematical simulations to justify my methodological analyses by arguing that linear mixed models provide the most suitable analytical approach to provide answers to the questions posed in this manuscript. I focus on presenting several statistical simulations and explore conditions under which results that are obviously significant for a linear mixed model might beget insignificant results for $F1$ and $F2$ analyses, and vice versa, by manipulating the effect of treatment in a variety of simulated datasets. Finally, I argue that the analyses used in this manuscript provide more accurate and reliable results than the standard statistical analyses used in the literature.

These chapters demonstrate that research on mental representations can benefit from an integrated viewpoint. I conclude in Chapter 7 by suggesting that it is less relevant for the cognitive sciences to consider whether conceptual processing is symbolic or embodied and it is instead important to determine when, why, and to what extent linguistic and perceptual representations are employed during language processing.

Chapter Two

Language statistics explain the spatial–numerical association of response codes

Abstract

The spatial–numerical association of response codes (SNARC) has shown that parity judgments with participants' left hands yield faster response times (RTs) for smaller numbers than for larger numbers, with the opposite result for right-hand responses. Participants perceptually simulating magnitude on a mental number line have explained these findings. In three RT experiments, the SNARC effect was also explained by language statistics. Participants made parity judgments of number words (Exp. 1) and Arabic numerals (Exp. 2). Linguistic frequencies of the number words and numbers mirrored the SNARC effect, explaining aspects of processing that a perceptual simulation account could not. Experiment 3 investigated whether high- and low-frequency nonnumeric words would also elicit a SNARC-like effect. Again, RTs were faster for high-frequency words for left-hand responses, with the opposite result for right-hand responses. These results demonstrate that what has only been attributed to perceptual simulation should also be attributed to language statistics.

This chapter is based on:

Hutchinson, S., & Louwerse, M. M. (2014). Language statistics explains spatial-numerical association of response codes. *Psychonomic Bulletin & Review, 21,* 470-478.

Hutchinson, S., Johnson, S., & Louwerse, M. M. (2011). A linguistic remark on SNARC: Language and perceptual processes in Spatial-Numerical Association. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual meeting of the Cognitive Science Society* (pp. 3437-3442). Austin, TX: Cognitive Science Society.

Introduction

In the previous chapter I provided an overview of the evidence supporting both embodied and symbolic representations. I also discussed the more recent emergence of integrated theories of cognition. Several psycholinguistic theories have found that experimental findings that have been attributed to perceptual simulation can alternatively be explained by language statistics (Louwerse, 2008; Louwerse & Jeuniaux, 2010). In the following chapter I will attempt to portray how linguistic theories of cognition can explain results for effects that are commonly attributed to embodied cognition. Just as cognitive scientists have argued that cognition is fundamentally embodied and that concepts are understood through perceptual simulation (Pecher & Zwaan, 2005; Semin & Smith, 2008), the same is argued for numerical information. Although intuitively number manipulation might seem more symbolic than perceptual in nature, as the computing of numbers does not require references to the symbols being manipulated or a visual representation of the manipulation process, a spatial representation of numbers is often thought to facilitate our understanding (Semenza, 2008). Evidence for this claim is plentiful, with participants processing small numbers faster with their left hand and large numbers faster with their right, akin to perceptually simulating numbers on a mental number line. This finding is known as the spatial–numerical association of response codes (SNARC; Dehaene, Bossini, & Giraux, 1993; Fischer & Brugger, 2011; Restle, 1970).

The SNARC effect is robust, with physical manipulations (e.g., crossing hands, grasping) and handedness failing to influence its direction (Andres, Ostry, Nicol, & Paus, 2008; Dehaene et al., 1993). SNARC holds for two-digit numbers (Dehaene et al., 1993; Reynvoet & Brysbaert, 1999) and number words (Fias, 2001), and extends to other ordinal-sequence-based organizational systems, such as alphabets, and large/small object words (Gevers, Reynvoet, & Fias, 2003; Ren, Nicholls, Ma, & Chen, 2011; Shaki & Gevers, 2011).

Several theories have been proposed to explain SNARC in terms of embodied cognition. Dehaene et al. (1993) suggested that numbers are spatially organized on a mental number line according to magnitude. Alternatively, SNARC might be an embodied association between numbers and actions (e.g., common patterns of motor activation are based on the left side of a keyboard having small numbers and the right side having large numbers (Gevers, Caessens, & Fias, 2005). Fischer and Brugger (2011) suggested that finger counting might be the origin of the effect. These theories share the idea that the SNARC effect is the consequence of embodied mechanisms.

Apart from simply expanding upon instances in which the SNARC effect occurs and does not occur, various theories explaining the effect have been offered. As number representation is thought to be somewhat independent of other language processes (Semenza, 2008), many researchers have proposed a spatial representation explanation of the SNARC effect. In other words, the SNARC effect occurs because the mental representations of numbers are spatially organized according to number magnitude (i.e., numbers are placed on a mental number line with small numbers on the left and large numbers on the right; Zorzi, Priftis, & Umilta, 2002). Although such an explanation is succinct and even empirically supported through neurological research (Zorzi et al., 2006), it fails to account for how numbers are represented for language users of specific groups that show reverse SNARC effects (Maass & Russo, 2003) or fail to show any SNARC effects (e.g., Israelis and illiterate Arabic speakers). Bächtold, Baumüller, and Brugger (1998) have posited that the SNARC effect might be due to a learned embodied association between numbers and actions (i.e., common patterns of motor activation make use of the knowledge that the left side of a keyboard possesses only small numbers whereas the right possesses large numbers). While Proctor and Cho (2006) claimed that the SNARC effect occurs through the consideration of stimuli polarity. According to a theory of number representation, small numbers have a negative polarity whereas large numbers have a positive polarity. Thus words and numbers are represented along a positive-negative dimension in space. In the instance of SNARC, the right side and large numbers are associated with a positive polarity and the opposite is true for the left side and small numbers. Even others suggest that two different processing routes (a top-down conditional route and an automatic unconditional route) work together simultaneously to help us understand the stimuli being presented, therefore accounting for RT differences among various numbers (Gevers, Cassens, & Fias, 2005; Gevers, Lammertyn, Notebaert, Verguts, & Fias, 2005). It is important to note that despite differences between theories, most agree the SNARC effect is, at least in some way, further evidence for perceptual simulation during cognition.

However, some findings have questioned a solely embodied account. For instance, the original task demonstrated vertical (Ito & Hatta, 2004) and horizontal (Shaki, Fischer, & Petrusic, 2009; Zebian, 2005) effects, indicating that the mental number line is not canonical. In addition, Arabic speakers show reverse SNARC effects reverse SNARC effects (Maass & Russo, 2003) and illiterate participants fail to show a SNARC effect at all (Zebian, 2005). Moreover, Fischer, Shaki, and Cruise (2009) found that spatial representation is not inherent in numbers, but caused by directional reading conventions. These findings suggest that embodied mechanisms might not be the only explanation for SNARC and hint at a linguistic explanation.

In line with Louwerse (2008), who argued that language is organized so that it reflects embodied relations, in this chapter I argue that numerical cognition can also be explained through both linguistic and embodied mechanisms. That is, the prelinguistic conceptual knowledge (e.g., number magnitude) used when speakers formulate utterances gets translated into linguistic conceptualizations (language statistics) so that, as a function of language use, embodied relations are encoded in language. The fact that findings originally attributed to embodied cognition can also be attributed to language statistics begets the question of whether SNARC might be attributed to statistical linguistic factors.

To test for this possibility, three experiments compared embodied and linguistic accounts as possible additional explanations for SNARC. Two experiments replicated the SNARC experiments with both number words (Dehaene et al., 1993; Fias, 2001; Nuerk, Iversen, & Willmes, 2004) and Arabic numbers (Dehaene et al., 1993). It was expected that there would be a strong negative correlation between number magnitude and number (word) frequency, as more frequent numbers are also smaller (e.g., 1 is more frequent than 2), with both accounts explaining response times (RTs). Furthermore, because language encodes embodied representations, a strong correlation between the perceptual ordering of the numbers and their frequencies was expected. A third experiment was conducted to investigate whether the collocation frequency of trial pairs could explain RTs, as this effect cannot readily be accounted for by embodied cognition. In general, evidence that word frequency elicits a SNARC effect would suggest that in addition to embodied representations, SNARC can also be explained by language statistics. In other words, if linguistic factors also explain the SNARC effect, the collocation frequencies of paired number words (e.g., *one* preceding *two*, *one* following *two*) should impact processing time.

Experiment 1: Number Words

In Experiment 1 I ask whether SNARC can be explained by number magnitude or by language statistics. As in most SNARC studies, participants were asked to evaluate whether numbers were even or odd, by responding using their left or right index finger. However, instead of presenting Arabic numerals, number words were presented instead. If the SNARC effect has a linguistic basis, it should at least be found with number words (cf. Fias, 2001).

Participants

A group of 57 right-handed native English-speaking undergraduate students participated for extra credit. Following Dehaene et al. (1993), in randomly assigned conditions participants were instructed to first respond to even numbers with their left hand and odd numbers with their right hand (n = 27), or to use the reverse mappings (n = 30).

Materials

Each experiment consisted of 65 trials, with each trial including two number words, ranging from one to nine (excluding five; Tzelgov, Meyer, & Henik, 1992). The rationale for using number words was that a) if the SNARC effect could have a linguistic basis, it should first and foremost be found in words and b) although making parity judgments regarding number words may seem to be more difficult than making parity judgments about Arabic numerals, still number words have shown to yield a SNARC effect (Fias, 2001).

Admittedly, there is evidence that number words and Arabic numerals are processed in different ways (Damian, 2004; Fias, 2001). However, past research has suggested that number word presentation shows few differences from traditional Arabic numeral presentation in a SNARC experiment (Nuerk, Iversen, & Willmes, 2004). Furthermore, as number words were exclusively presented, any variations in RTs should be systematic across all parity judgments, and are thus of little consequence.

Procedure

Number words were presented in the center of an 800 × 600 screen in 36-pt. font and subtended at most 2.5o of vertical visual angle from 60 cm. Two words were presented in each trial, but the words appeared on the screen one at a time. Participants were asked to determine number parity. After a participant responded to the first word, the second was presented. Although participants responded to all of the words, only RTs to the second word in each trial were analyzed. The stimuli within each trial were paired so that participants saw each number paired with every other number, in both orders (e.g., participants saw trials of both one followed by three and three followed by one). This allowed for all word pair frequencies to be accounted for. Once a participant had responded to both words in a trial, the next trial would commence after a short beep and after the "+" symbol had appeared for 1,000 ms. Every trial was so separated as to provide space between the trials. Trial pairs were randomly presented, and participants saw every combination of pairs. Six practice trials preceded the experiment.

Results and Discussion

Five participants were removed because >14% of their answers were incorrect. This threshold was selected as the natural cut-off after visual inspection of error rates. After removing those five participants, the average error rate was 5%. Outliers were identified as responses faster than 200 ms or slower than 1,500 ms, following the criteria of Shaki et al. (2009). Errors and outliers were removed, affecting 6.5% of the data.

As in Dehaene et al. (1993) and Fias (2001), the median RT per number word per response side was separately computed per participant.[1] Median left-hand responses were subtracted from median right-hand responses. A mixed-effects regression was conducted on RTs, with response side and magnitude as fixed predictors and participant and item as random predictors (Baayen et al., 2008; Hutchinson, Wei, & Louwerse, 2014) to predict whether a SNARC effect was replicated. The model was fitted using restricted maximum likelihood estimation (REML) for the continuous variable (RT). The F-test denominator degrees of freedom were estimated using Kenward–Roger's degrees-of-freedom adjustment, in order to reduce the chance of Type I error (Littel, Stroup, & Freund, 2002). Evidence supporting an embodiment account stems from the interaction between faster left-hand responses for smaller numbers and faster right-hand responses for larger numbers, as this interaction links space (right/ left hand) and magnitude. A main effect emerged for response side, with faster RTs for right-hand responses, F (1, 5815.85) = 6.57, p = .01, R^2= .10. This result is not surprising, as all participants were right-handed. More interestingly, there was an interaction between response side and magnitude, F (1, 5816.93) = 3.26, p = .04, R^2= .04 (Figure 1), replicating the SNARC effect and providing support for an embodied explanation of the effect.

A second regression with response side and linguistic frequency as fixed predictors was also performed to determine if linguistic factors could also explain the effect. The linguistic factor was operationalized as the log frequency of the number word (see Table 2). This value indicates how frequently each number appears in a large corpus. Specifically, word frequencies were obtained from the Web IT one-trillion-word 5-gram corpus (one trillion word tokens, with 13,588,391 word types from 95,119,665,584 sentences; Brants & Franz, 2006). Log frequency is typically preferred over raw frequency because the distribution of word frequency is right-skewed (i.e., L-shape; Baayen, 2001).

As predicted, there was a strong negative correlation between magnitude and word frequency, $r = -.98$, $p < .001$ (cf. Dehaene & Mehler, 1992). In other words, the more frequent a number word, the lower its magnitude (i.e., *one* is more frequent than *nine*). This allows for the possibility that SNARC can also be explained by word frequencies. The SNARC effect traditionally predicts small numbers to be processed faster with the left hand; if word frequency alone affected RTs, then faster processing of frequent words would be expected regardless of response side. Therefore, if linguistic frequency plays a role during numerical processing, frequency should then not affect RTs, but an interaction should exist between response side and frequency. As expected, frequency did not explain the RTs, F (1, 5587.95) = .01, p = .93, R^2= .0003, but, analogous to the SNARC effect, an interaction was apparent between response side and frequency, F (1, 5586.16) = 3.23, p = .04, R^2= .04 (Figure 2) meaning frequent words were processed faster with the left hand, and less frequent words were processed faster with the right hand.

Whether the linguistic system simply provides redundant information derived from the perceptual system is still unanswered, because what is

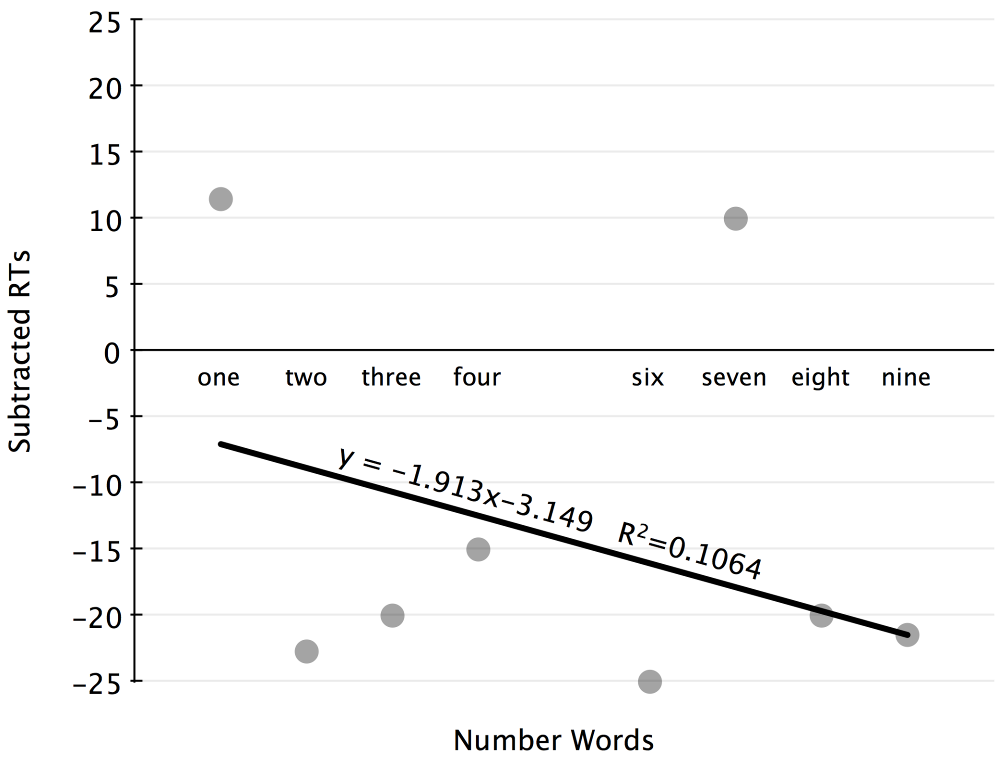


Figure 1. Linear fitting of the SNARC effect for Experiment 1

explained by word frequency is also explained by number magnitude. To test whether linguistic frequencies independently explain the findings, the collocation frequencies of paired number words in each trial were analyzed (see Table 2). If statistical linguistic frequencies of the word pairs explain RTs, this finding would be difficult to attribute to embodied mechanisms because collocation frequencies cannot be explained by the magnitude of the second word. No correlation emerged between collocation frequencies and the second number's magnitude, $r = -.15$, $p = .20$. In a mixed-effects model, bigram frequency significantly explained RTs of the second word in each pair, F (1, 3072.72) = 4.12, $p = .04$, $R^2 = .14$, with higher frequencies yielding lower RTs. A significant interaction was found between response side and frequency, F (2, 3082.32) = 3.54, $p = .03$, $R^2 = .12$. These collocation results thus mirror the

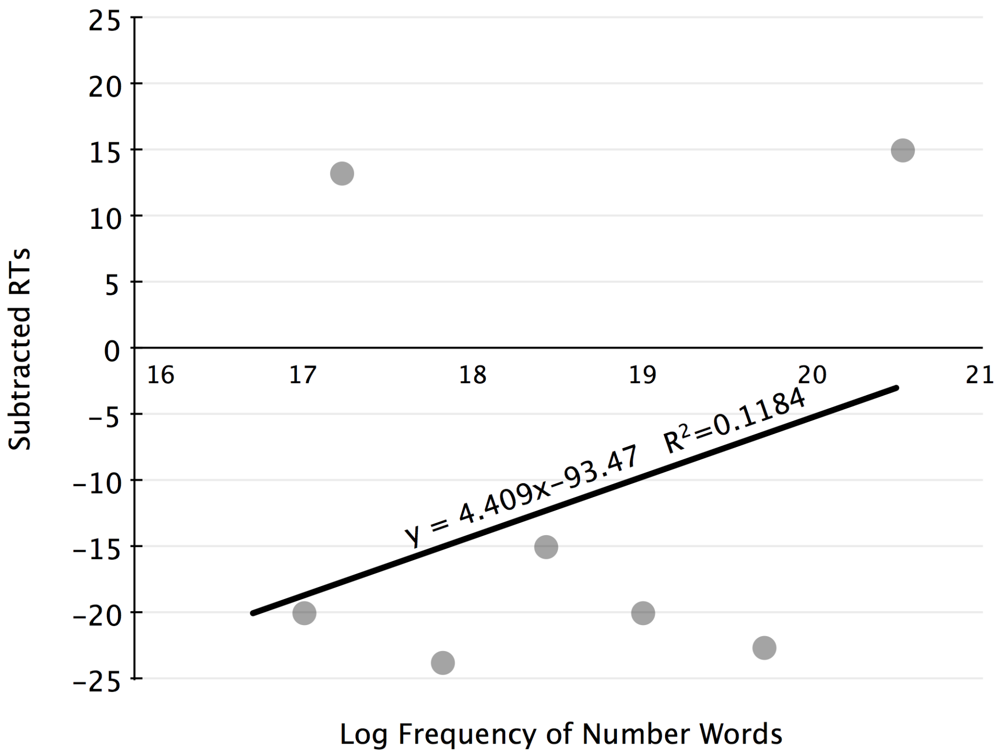


Figure 2. Linear fitting of the statistical linguistic frequencies for Experiment 1

Table 1

Results from all three experiments

	Exp. 1		Exp.2		Exp. 3	
	df	*F*	*df*	*F*	*df*	*F*
Magnitude	5817	1.01	4973	0.10		
Frequency	5588	0.18	4973	2.37	1856	3.24
Response Side x Magnitude	5817	12.52**	4973	13.88**		
Frequency x Magnitude	5586	8.67**	4973	14.60**	1856	7.33**
Bigram Frequency x Response Side	3082	13.77**	2098	18.34**		

** $p < .01$

traditional SNARC findings, but they are difficult to explain with an embodied account, because the magnitude of the second word does not correlate with the frequency with which the two words appear together, providing evidence for an independent linguistic account.

Experiment 1 demonstrated that language statistics explain RTs as well as an embodied cognition account, suggesting that indeed those effects attributed to embodied perceptual representations might also be explained through linguistic representations. However, the argument could be made that Experiment 1 used number words and therefore was biased toward a linguistic account. Furthermore, these significant effects resulted in low effect sizes, therefore two additional experiments were conducted to verify and expand the results.

Table 2

Bigram and unigram log frequencies

	Unigram	Bigram							
		one	two	three	four	six	seven	eight	nine
one	20.72	16.54	15.36	14.29	13.38	12.90	12.12	11.80	11.66
two	19.91	16.45	14.73	13.80	13.25	12.37	11.58	11.48	11.02
three	19.22	15.41	15.91	13.87	12.91	12.28	11.87	11.18	11.20
four	18.68	14.73	14.45	15.28	13.58	11.99	11.26	11.49	10.71
six	18.08	13.80	13.02	13.53	13.75	12.91	11.44	11.25	10.77
seven	17.61	13.41	12.25	12.45	12.45	13.61	13.09	11.09	10.75
eight	17.30	13.23	12.09	11.88	12.86	13.45	13.22	11.72	10.80
nine	17.10	12.78	11.43	12.24	11.37	12.46	12.40	12.81	12.76

Experiment 2: Arabic Numerals

In Experiment 2, Arabic numerals were used instead of number words, as Arabic numerals may be processed differently from number words (Damian, 2004). The number 0 was also included, whose low magnitude, yet lower frequency than other low-magnitude digits, allowed for comparing an embodied account (that magnitude explains SNARC) and a frequency account (that frequency explains SNARC) (cf. Pinhas & Tzelgov, 2012). In other words, the number 0 appears less frequently than the number 1, yet its magnitude is less than 1.

Participants

A group of 44 right-handed native English-speaking undergraduates participated for extra credit. The participants were evenly split between response conditions.

Materials

Each experiment had 81 trials, including two Arabic numerals presented one at a time, ranging from 0 to 9 (excluding 5).

Procedure

The procedure, font size, and viewing angle were identical to those in Experiment 1. Participants were again asked to determine number parity, with instructions that specified that 0 was an even number.

Results and Discussion

Eight participants were removed because >14% of their answers were incorrect. A software error led to the loss of 2.2% of the data. Outlier removal resulted in data loss of 2.43%.

The analysis was the same as in Experiment 1, in which median RTs per number word per response side were separately computed for each participant. Response side explained RTs, $F(1, 4973) = 1.20$, $p < .001$, $R^2 = .02$, and the interaction between response side and magnitude was significant, $F(1, 4973) = 13.88$, $p < .001$, $R^2 = .23$ (Figure 3), replicating the SNARC effect.

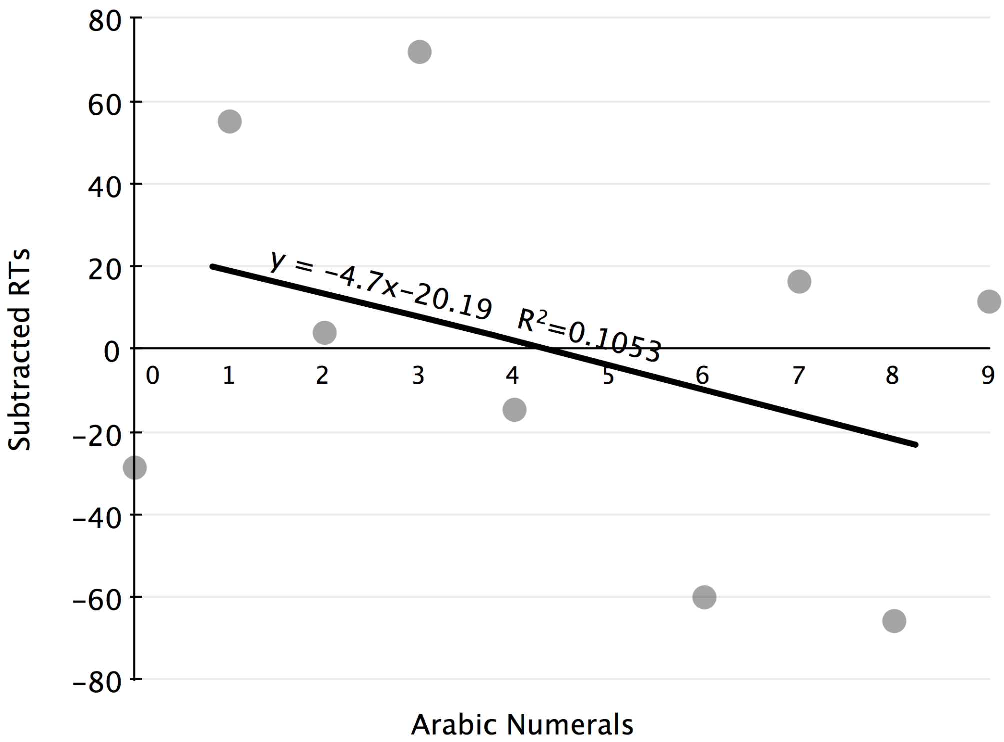


Figure 3. Linear fitting of the SNARC effect for Experiment 2

Similar to the negative correlation between number words and magnitude in Experiment 1, there was a negative correlation between Arabic numerals and their frequencies, $r = -.60$, $p < .001$. Note that the correlation was weaker than before, because of the inclusion of 0. Without 0, the correlation was stronger, $r = -.98$, p < .001. Frequency did not affect RTs, $F(1, 4973) = 0.05$, $p = .81$, $R^2 = .001$, but the Response Side × Frequency interaction was significant, $F(1, 4973) = 14.60$, $p < .001$, $R^2 = .24$ (Figure 4). This finding replicated the SNARC effect and is similar to the results of Experiment 1, except that it was now obtained with numbers rather than number words.

As before, frequency collocations for pairs were assessed to determine whether bigram frequency alone impacted RTs. Bigram frequencies did not

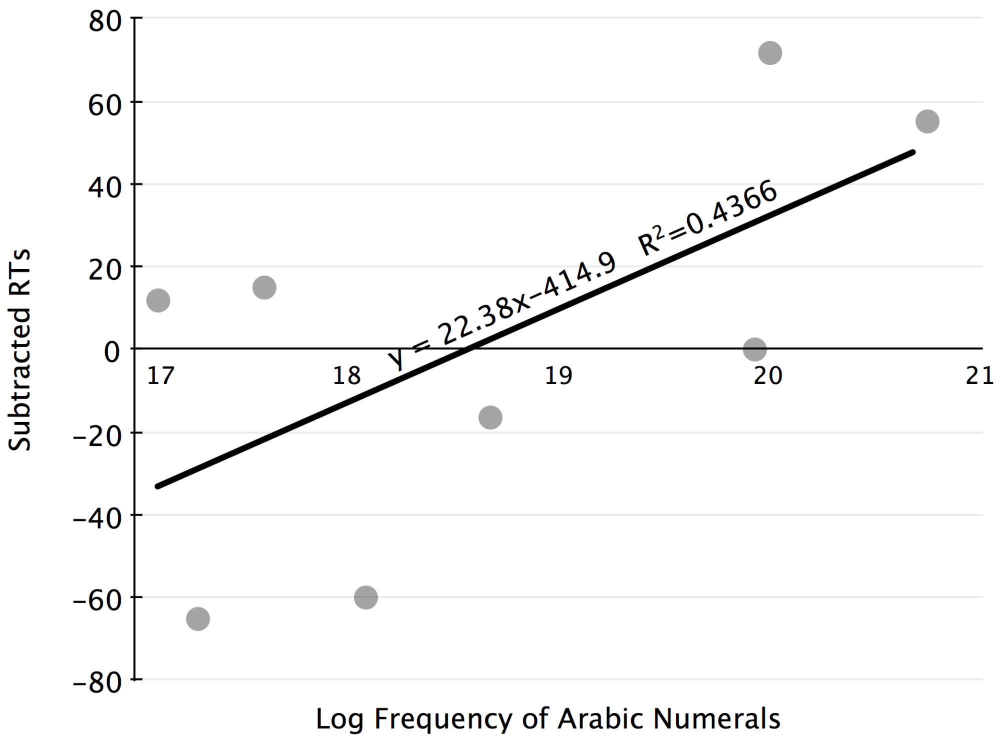


Figure 4. Linear fitting of the statistical linguistic frequencies for Experiment 2

correlate with the magnitude of the second word in the pair, r = .08. A main effect of response side was found, $F(1, 2098) = 9.29$, $p < .01$, $R^2 = .12$, and bigram frequency did not significantly explain RTs, $F(1, 2098) = .03$, $p = .88$, $R^2 = .001$. Importantly, the Response Side × Frequency interaction was significant, $F(1, 2098) = 42.22$, $p < .001$, $R^2 = .53$, a finding that cannot be explained by an embodied account. See Table 3 for the bigram and unigram log frequencies of the Arabic numerals.

Including 0 allowed for a comparison of the two accounts, because 0 has the lowest mathematical and psychological magnitude (Pinhas & Tzelgov, 2012), yet it has a lower frequency than the other low-magnitude numbers. Left-hand responses for 0 were slower (M = 670 ms) than right-hand responses (M =

Table 3

Bigram and unigram log frequencies

	Unigram	Bigram								
		0	1	2	3	4	6	7	8	9
0	21.34	21.06	19.24	18.15	17.64	17.30	16.77	16.61	16.66	16.24
1	21.33	18.89	19.68	18.21	17.71	17.22	16.72	16.60	16.46	16.09
2	21.66	17.97	19.72	18.38	17.53	17.18	16.49	16.36	16.16	15.88
3	20.99	17.47	19.29	18.97	17.91	17.03	16.30	16.06	15.86	15.64
4	20.77	17.23	18.68	18.66	18.81	17.50	16.43	15.99	16.03	15.57
6	20.22	16.73	17.12	17.88	17.93	18.24	17.02	16.21	16.48	15.73
7	20.19	16.27	16.91	16.10	17.75	17.80	18.20	17.26	16.14	15.82
8	19.97	16.53	17.02	16.15	16.25	17.71	17.95	18.05	16.95	16.03
9	17.08	16.10	16.49	15.81	15.82	15.65	17.63	17.82	17.94	17.06

641 ms), albeit not significantly, $t(555.65) = -1.5$, $p = .13$. To determine whether the RT findings for 0 provided support for a frequency or embodied account, RTs for the items 0 and 1 were compared. If magnitude explained responses, because both numbers shared low magnitudes, no significant difference was expected between them. But if word frequency explained the responses, because 1 is quite frequent and 0 is less frequent, the RTs for these two items were predicted to be divergent, which was what was found, $t(10.75) = -4.5$, $p < .001$. However, the differences between 0 and 1 might be explained by a learned embodied relationship, with the “0” key on a keyboard being on the right. To support such an explanation, it would be necessary for RTs to 0 to be faster with the right hand, but they were not, $t(555.65) = -1.5$, $p = .13$.

Experiments 1 and 2 demonstrated that a language statistics account could also offer an explanation for the SNARC effect. This evidence does not replace the embodied SNARC effect, because there was an effect of magnitude in both experiments. In fact, in Experiments 1 and 2, participants seemed to use frequency information when making parity judgments about either number words (Exp. 1) or Arabic numerals (Exp. 2), as was evidenced by the significant Magnitude × Response Side and Frequency × Response Side interactions. Although Damian (2004) claimed that number words and Arabic numerals are processed differently, such that when processing Arabic numerals information about magnitude is more readily available, and when processing number words, lexical information is more readily available, participants in Experiments 1 and 2 were asked to make parity judgments, calling explicit attention to neither magnitude nor frequency.

With the findings from Experiments 1 and 2, at least two arguments bolster a complimentary linguistic frequency explanation for the effect. First, when 0 was included—a number with both low magnitude and lower frequency—no SNARC effect was found, even though a frequency effect was obtained. Second, bigram frequencies explained the RTs, whereas such an explanation is lacking for an embodied account. However, a problem with a language statistics account concerns direction. Whereas number-line representations explain why left-hand responses are faster for low-magnitude items, the rationale is not so obvious for high-frequency numbers eliciting faster left-hand responses.

Markedness can explain this pattern. Greenberg (1966) argued that for any word pair, the one that is more frequent is the unmarked (i.e., most natural, simplest, first learned), and the one that is less frequent is the marked member of the pair, with unmarked members preceding marked members (Louwerse, 2008). Although this explanation seems similar to Proctor and Cho's (2006) polarity correspondence principle, a markedness explanation suggests that for any given pair, items will be processed faster when frequent items appear before infrequent items, not when items are matched (to their response sides) on polarity. This general linguistic theory of markedness ranges over phonological, grammatical, and semantic elements (Chomsky & Halle, 1968) and can be applied to number words, with frequent items being processed faster with the left hand. The bigram frequencies for pairs of number words in the Web IT one-trillion-word 5-gram corpus (Brants & Franz, 2006) show that frequent number words precede infrequent number words more often than vice versa, $F(1, 70) = 31.25$, $p < .001$. If the frequency asymmetry is an explanation for the direction of the language statistics effect found in Experiments 1 and 2, then a "SNARC" effect is expected, with high-frequency (non-magnitude) words being processed faster with the left hand, and vice versa for the right. This hypothesis was tested in Experiment 3.

Experiment 3: High and low Frequency Words

Participants

A group of 49 students participated for extra credit. Of these participants, 22 were randomly assigned to first respond to animate words with their left hand and inanimate words with their right hand, and 27 to the opposite mapping.

Materials

In all, 30 two-word trials were presented one word at a time. The words extracted from the MRC Psycholinguistic Database were frequent or infrequent and were matched on word length (see Table 3). The word frequencies of frequent and infrequent words differed significantly, $t(69) = -17.10$, $p < .001$. Half of the words described animate concepts, whereas the rest described inanimate concepts.

Procedure

The procedure, size, and viewing angle were identical to those in Experiments 1 and 2. Participants were asked to indicate whether a word that appeared on the screen represented something animate or inanimate.

Results and Discussion

Seven participants were removed because >14% of their answers were incorrect. Outlier removal resulted in the loss of 2.1% of the data. Median RTs

Table 4

Unigram log frequencies of the experimental stimuli for Experiment 3

Inanimate Words	Log Frequency	Animate Words	Log Frequency
bank	18.36	ant	15.77
bin	18.75	bat	16.07
carpet	16.43	bear	17.56
chilly	14.70	boy	18.26
cord	16.78	bunny	15.45
cut	18.60	camel	16.92
gold	18.19	child	19.18
groan	13.94	deer	16.25
gun	17.39	dog	18.51
happy	18.45	dove	14.84
ivory	15.21	fish	18.20
law	19.49	flower	17.41
marry	16.13	girl	18.89
near	1.03	hog	14.90
numb	14.66	horse	18.37
number	20.37	human	19.31
ocean	17.10	insect	15.77
plane	17.45	lynx	15.00
pump	17.17	man	19.62
rake	14.86	moose	14.65
rim	15.63	person	19.69
ship	18.55	pigeon	14.60
shoe	16.87	plant	18.30
sleep	17.82	puppy	16.07
stove	15.70	rose	17.41
summer	18.34	seal	16.65
text	19.74	sheep	16.53
under	20.40	tree	18.34
vigil	14.20	woman	18.75
zoo	16.40	worm	16.11

per word per response side were separately computed for each participant (see also note 1). The same mixed-effects regression was conducted on RTs as in the previous experiments. Response side significantly predicted RTs, $F(1, 1856) = 3.73$, $p = .05$, $R^2 = .14$, with right-side responses being faster. Frequency approached significance, $F(1, 1856) = 3.24$, $p = .07$, $R^2 = .12$. Importantly, the Response Side × Frequency interaction was significant, $F(1, 1856) = 7.23$, $p < .01$, $R^2 = .27$, indicating that high-frequency words were indeed processed faster with the left hand, whereas low-frequency words were processed faster with the right (Figure 5).

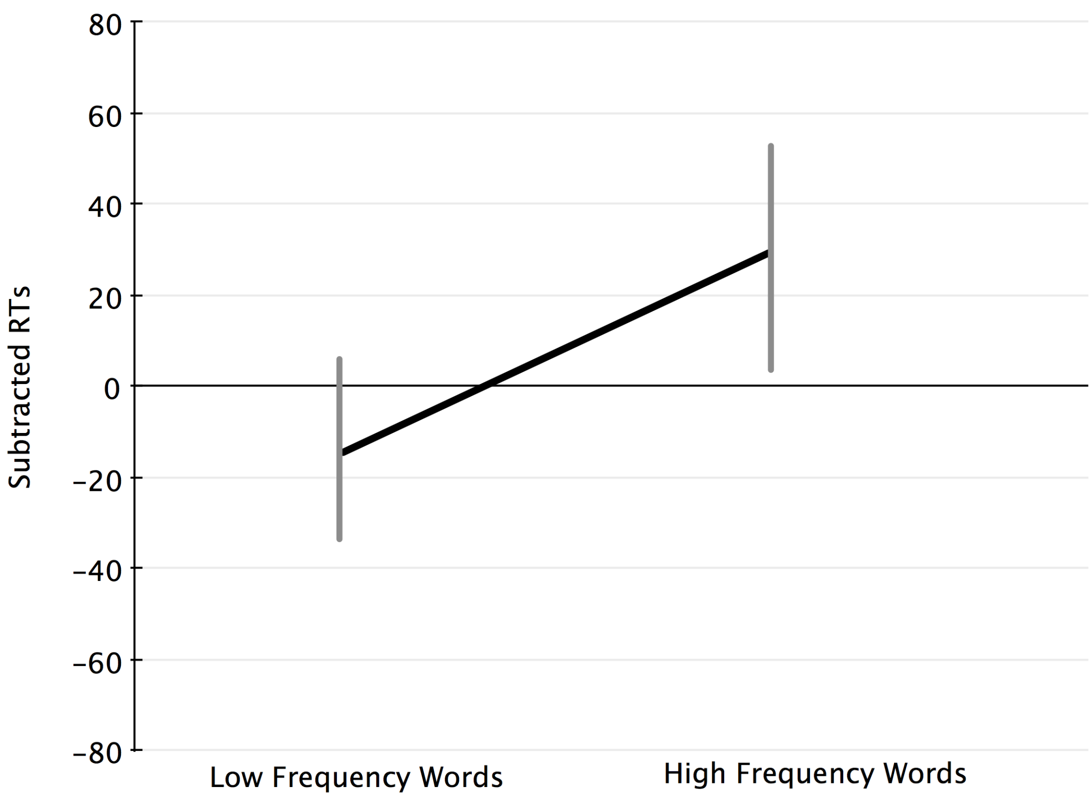


Figure 5. Linear fitting of the statistical linguistic frequencies for Experiment 3

Conclusion

In the previous pages, I have empirically demonstrated in three experiments that both linguistic and perceptual information is relevant during conceptual processing. The first and second experiments focus on numerical stimuli, replicating the well-known embodied effect known as the SNARC effect and offering a complimentary linguistic explanation for this effect. The third and final experiment demonstrates that responses to linguistic stimuli also generate results that can be explained by both perceptual and linguistic accounts. The SNARC effect has traditionally added to the large body of literature introduced earlier that suggests that cognition is fundamentally embodied. Yet several studies have demonstrated that language statistics can explain the experimental findings equally well (Louwerse, 2008; Louwerse & Jeuniaux, 2010). After an examination of the use of linguistic and perceptual mental representations for numerical and linguistic stimuli, I provided evidence that both types of representations work together to establish meaning, since both embodied and linguistic factors explain participant response times from three experiments, two of which the effects were traditionally accounted for by perceptual representations. Furthermore, linguistic collocation frequencies were able to explain response times to experimental tasks where an embodiment account could not. In addition, response times from the number 0, a number with both low magnitude and frequency, supported a linguistic account and not an embodiment account, suggesting that linguistic representations certainly play a role when processing numeric stimuli. Finally, a SNARC-like effect was found for low-frequency and high-frequency words, for which embodiment could not be the explanation.

The finding that linguistic frequencies explain the SNARC does not dismiss an embodiment account. After all, the interaction of response side and word frequency is considered and accounted for by embodied representations. However, the source of SNARC is not necessarily magnitude on a perceptually simulated mental number line. Perhaps language has encoded such perceptual number line information, so that language users rely on language statistics in during their cognitive processes (Louwerse, 2011). Consequently, frequency would then be likely to explain SNARC-like effects obtained with a variety of stimuli, whether magnitude information was present or not, such as with ordinal information or even large/small object words (Gevers et al., 2003; Ren et al., 2011; Shaki & Gevers, 2011).

The notion of frequencies playing a role in numerical cognition is not new. Dehaene et al. (1993) evaluated the interactions between number and word representations and showed that treating them as eliciting separate processes is not an accurate description of number processing. This conclusion is reminiscent of the conclusion drawn by Louwerse (2008) that the nature of conceptual processing is symbolic and embodied. Language statistics facilitate cognitive processes because language encodes magnitude. Such findings demonstrate that indeed, integrated theories can more easily account for effects together than each theory can independently.

Having established that both linguistic and perceptual representations are important during conceptual processing, I will next address how different factors can impact how much participants rely on each of these types of representations, by breaking down the sequence of an experiment to consider whether the time course and spatial presentation of stimuli might impact the results.

Footnotes

[1] The same analysis, but using the mean RT per number word (or numeral) per response side per participant, so that mean (rather than median) left-hand responses were subtracted from the mean right-hand responses, yielded similar results for all three experiments (see Table 1).

Chapter Three

Time, space, and independence

Abstract

In three Experiments, the factors of time and space, as well as the independence of symbolic and embodied representations were more thoroughly explored. Experiment 1 investigated whether time constraints impacted the use of perceptual and linguistic factors during language processing. Participants made fast or slow semantic judgments about pairs of words. A linguistic factor best explained fast RTs but when given more time to respond, both linguistic and perceptual representations were used. Experiment 2 explored absolute or relative location of concept-location words and whether semantic judgments were made with respect to an absolute location on the screen (an embodied explanation) or with respect to a relative location in comparison to other words included in the experimental session (a statistical linguistic explanation). In a response time experiment, there was a concept location facilitation effect for words presented at various locations on the screen, supporting the view that language processing is both linguistic and embodied. In Experiment 3, I demonstrated symbolic and embodied processes are indeed independent. Relying on the finding that linguistic processing is more important in the early stages of language processing and perceptual processing is more important later, participants were asked to determine whether linguistically and/or perceptually similar or dissimilar word pairs are semantically related. For unrelated word pairs, perceptual factors alone predict RT performance, suggesting that a full perceptual representation is independently utilized when generating a relationship for unrelated word pairs.

Together these findings show that language processing is both linguistic and embodied, but to different extents in different situations.

This chapter is based on:

Hutchinson, S., Tillman, R., & Louwerse, M. M. (2015). *Relating the unrelated*. Manuscript in preparation.

Hutchinson, S., & Louwerse, M. M. (2013). What's up can be explained by language statistics. In M. Knauff, M. Pauen, N. Sebanz, & I. Washsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 2596-2601). Austin, TX: Cognitive Science Society.

Hutchinson, S., Tillman, R., & Louwerse, M. M. (2014). Quick linguistic representations and precise perceptual representations: Language statistics and perceptual simulations under time constraints. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2399-2404). Austin, TX: Cognitive Science Society.

Introduction

Both linguistic and perceptual properties are relied upon during processing – a finding that I referred to in the previous two chapters. I further confirmed and replicated this finding in the current chapter. Furthermore, I asked what kind of conditions encourage participants to rely more or less on embodied or linguistic representations, and how we know these processes are independent. Specifically, I explored the potential that temporal and spatial constraints on stimuli presentation might influence linguistic and embodied mechanisms. Both of these aspects are important because a) language processing is time constrained and b) much of the evidence for embodied representations comes from word on words sharing spatial relationships. First I examined how the use of linguistic and perceptual representations were impacted when the time course of an experimental trial was constrained. I then explored how the spatial presentation of stimuli on the screen also impacted how and when participants were more or less likely to rely on linguistic versus perceptual representations. Finally, in a third and final experiment, relying on the findings from the first experiment, I demonstrated that both linguistic and perceptual representations, although intertwined are still independent processes.

Embodied cognition studies have demonstrated that when words that are typically associated with higher physical locations (e.g., *bird*) are positioned at the top of a screen they are processed faster than when they are positioned at the bottom of the screen. The reverse effect is obtained for words typically associated with lower physical locations (e.g., *fish*). This concept-location facilitation effect has been argued to demonstrate that cognitive processing is fundamentally perceptual in nature. As demonstrated in the prior chapters, there is increasing evidence in the past several decades from experiments that language statistics and perceptual simulations both play a role in conceptual processing. These studies demonstrate that both language statistics and perceptual simulation must be taken into consideration together. For instance, the relative importance of language statistics and perceptual simulation in conceptual processing depends on several variables, including factors like the type of stimulus presented to a participant, or the cognitive task the participant is asked to perform (Louwerse & Jeuniaux, 2010). In the following two experiments, I demonstrate that constraints on time and concept location also impact the relevance of each type of representation.

Evidence supporting only an embodied cognition account comes from single words, presented in different locations on a computer screen. For example, Šetić and Domijan (2007) presented 'up' and 'down' words one at a time either in a perceptually expected location (e.g., *butterfly* appeared at the top of the screen) or an unexpected location (e.g., *butterfly* appeared at the bottom of the screen). Participants were asked to determine if the word they saw was something animate (living animal) or something inanimate (non-living entity). As expected, participants were faster to process concept-location matches (e.g., the *butterfly* presented at the top of the screen) than concept-location mismatches (e.g., the *butterfly* presented at the bottom of the screen). Šetić and Domijan argued that this effect occurred because each word was associated with perceptual properties relevant to that object, like its location in space relative to an observer. In fact, Šetić and Domijan summarize that the spatial registration hypothesis (Coslett, 1999) should apply to all spatial directions, such that the absolute spatial location of a word on a screen should result in faster processing when the position of the word matches the real world position of the object described (e.g., left, right, top, bottom). Zwaan and Yaxley (2003) provide support for such a claim, by demonstrating that 'up' and 'down' words show no processing advantages when presented to the left and right of one another.

Unlike experiments comparing word pairs, findings for words in isolation, such as those in Šetić and Domijan (2007), are more difficult to also explain with a statistical linguistic account. That is, unigram word frequency does not explain congruency effects, as the set of 'up words' are not all more or less frequent than the set of 'down words'. In fact, when comparing how frequently the 'up words' and 'down words' occurred in a massive corpus of the English language (the *Web 1T 5-gram* corpus; Brants & Franz, 2006), no difference was obtained between the frequencies of 'up words' and 'down words' , $t(153.37) = 0.64$, $p = .52$. Consequently, the concept-location word results only seem to support an embodied cognition account and are argued to be due to the congruency of the presentation location and the perceptual features of the word: *Butterfly* is processed quickly at the top of the screen because a mental simulation of a butterfly involves perceptual and spatial information about where a butterfly is found in the actual world (above the ground/at the top). This poses a challenge to an account that argues for both linguistic and perceptual simulations factors in conceptual processing, such as proposed by the Symbol Interdependency Hypothesis (Louwerse 2007).

Although it seems straightforward to conclude that these effects must be due to the mental simulation of words, there are alternative explanations. Lakens (2011a; 2011b) argues that such effects might instead be due to polarity correspondence. Proctor and Cho (2006) found that in binary classification tasks, concepts could be processed faster when their polarity matches the response polarity. In other words, when a stimulus and a response are coded as either both positive or both negative, processing is facilitated, e.g., *butterfly* is processed quickly at the top of the screen because its location is positive (up), as is the response to whether or not it is found in the sky (yes). In order to rule out a polarity correspondence explanation for the results, in a similar experiment, Pecher, van Dantzig, Boot, Zanzolie, and Huber (2010) asked participants to respond to the question *Is it usually found in the ocean?* or to the question *Is it usually found in the sky?* They argued that for a polarity correspondence explanation to be valid, *yes* responses should be processed faster at the top of the screen, regardless of the question being asked, and regardless of word meaning. For instance, when being asked if an animal is found in the ocean, one would expect *butterfly* to be processed faster at the bottom of the screen because it is not found in the ocean, a hypothesis contrary to an embodied cognition explanation and a hypothesis that was not supported. Instead, the results showed just the opposite, i.e., when being asked if the animal is found in the ocean, *butterfly* was still processed faster at the top of the screen. In a response, Lakens (2011b) still suggested that perhaps *butterfly* is processed faster at the top of the screen, even when participants are making an ocean judgment because the judgment becomes a relative assessment with down as the default response (as all comparisons are made with reference to the ocean, which is down).

Lakens (2011b) goes further to point out that alternative explanations for data explained solely by perceptual simulations should not be overlooked. In addition, Lakens (2011b) and Louwerse (2011b) both suggest that results from Pecher et al. (2010) might likely also be explained by a statistical linguistic account. That is, although Pecher et al. (2010) concludes that mental simulation accounts for responses in the sky/ocean task, linguistic frequencies do contribute to word meaning and should also be considered. To illustrate, Louwerse (2011a) found that ocean animal names paired with the word *ocean* occur more frequently than ocean animal names paired with the word *sky* (and vice versa for sky animal names) and that these frequencies account for participant RTs. These findings illustrate that task instructions might influence response times because *ocean* and *sky* are more or less linguistically associated with the stimuli. In other words, linguistic information also contributes to our processing of these words, making it clear that perceptual simulations are more or less relevant in different scenarios.

For example, Louwerse and Jeuniaux (2010) found that both task and stimulus influenced whether participants were more likely to rely on linguistic or perceptual information. When participants were asked to make a judgment about word pairs, the statistical linguistic frequency of the word pair best predicted RTs whereas when participants were asked to make a judgment about image pairs, perceptual ratings about the pair better accounted for RTs (Louwerse & Jeuniaux, 2010). The cognitive task is also a particularly relevant factor when determining the dominance of linguistic or embodied representations, as one could imagine that for different scenarios, the relative salience of a representation may become more or less important. For example, Louwerse and Jeuniaux (2010) found that participants depended more on visual perceptual information when making iconicity judgments and more on linguistic information when making semantic judgments. Although both the linguistic and perceptual information about the word pair showed to be relevant in both cognitive tasks, with both verbal and non-verbal stimuli, different types of information were more, or less, important across different conditions. Because the iconicity judgments imply semantic judgments, Louwerse and Jeuniaux argued that language statistics best explain shallow, good-enough representations, and perceptual simulation best explain detailed full-fledged representations. This hypothesis does not reject either account but rather suggests that both perceptual and linguistic representations are used during processing and that their prevalence is relative depending on the cognitive task at hand.

Other factors impact linguistic and embodied representations as well. Borghi et al. (2004) found that when participants used their entire arm to respond to stimuli, an embodied effect is stronger than when only pressing buttons with their fingers. Individual differences in skill (Madden & Zwaan, 2006) and age (Dijkstra, Yaxley, Madden & Zwaan, 2004) have also shown to impact the strength of embodied effects.

We also know from previous research that in adults the relative prominence of linguistic representations precedes that of perceptual representations. Importantly, these studies do not deny or reject the importance of perceptual processes. After all, the Symbol Interdependency Hypothesis argues that language encodes perceptual information, making it difficult to disentangle the two variables. That is, effects attributed to statistical linguistic frequencies could also be attributed to perceptual simulation and vice versa. In fact, Louwerse and Connell (2011) found evidence for individual effects for perceptual simulations being seen early on in a trial when comparing the effect sizes of language statistics and perceptual simulations. Louwerse and Connell (2011) found that when comparing the effect of language statistics and perceptual information on response times, language statistics best explained quick response times, whereas perceptual information best explained slow response times (with both language statistics and perceptual simulation equally contributing to medium response times), with fuzzy regularities in linguistic context being used for quick decisions and precise perceptual simulations being used for slower decisions. Louwerse and Hutchinson (2012) extended the Louwerse and Connell (2011) study in an EEG experiment, demonstrating that linguistic cortical regions were relatively more active early in a trial and perceptual cortical regions were relatively more active later in a trial. And although processing is both symbolic and embodied throughout the trial, processes related to symbolic cognition (i.e., linguistic frequency) are relatively more prominent in the early stages of comprehension whereas processes related to grounded cognition are relatively more prominent later in processing. As participants glean just enough information from word co-occurrences to understand the relationship between word pairs immediately, while perceptual representations then take over to fill in the rest of the picture.

These studies demonstrate that both language statistics and perceptual simulation must be taken into consideration together, as both factors contribute to language processing. After all, the Symbol Interdependency Hypothesis argues that language encodes perceptual information, making it difficult to disentangle the two variables. That is, effects attributed to statistical linguistic frequencies could also be attributed to perceptual simulation and vice versa. If a more detailed representation is required, perceptual simulations allow for a precise mental representation. In essence, language acts as a sort of shortcut for language users by encoding symbolic and embodied relations in the world.

In this chapter, I will explore how linguistic and perceptual representations are employed differently under different conditions. In order to do so I will focus on two aspects: time and space. That is, I first focus on an experiment in which participants made fast or slow speeded judgments about whether pairs of words were semantically related. Subjects were also instructed to either respond as quickly as possible to the words they were presented, or respond as accurately as possible. In a second experiment, semantic judgments about individually presented words were made with respect to an absolute location on the screen (embodied explanation) or with respect to a relative location in comparison to other words included in the experimental session (statistical linguistic explanation). In other words, this chapter is not focused on exhaustively pinpointing how and when different types of mental representations are used. Rather, the question I explore is how and when temporal and spatial constraints in specific influence linguistic and perceptual representations. In particular, to explore these factors, two experiments compared embodied and linguistic accounts as possible explanations for RTs during two different semantic judgment tasks.

Experiment 1: Time

Although it is quite clear that linguistic representations are more prominent early in processing, and perceptual representations are more prominent later, studies that demonstrate that the relative prominence of linguistic representations precede that of perceptual representations, leave the question open whether RTs actually decrease because perceptual representations are facilitating processing or rather because the task duration biases participants to employ embodied representations. In most, if not all, embodied cognition experiments, participants have no pressing time constraints on their responses during a semantic judgment task, other than perhaps several seconds in order to force a decision. It is therefore possible that such a strong effect of perceptual simulation occurs during word processing because participants are allotted a longer time to process the words they are seeing. Put simply, is it because we are making a slower decision that we rely upon perceptual processes, or is it because we are relying upon perceptual processes do we then make a slower decision? Likewise, is it because we are making a faster decision that we rely upon linguistic processes, or is it because we are relying upon linguistic processes do we then make a faster decision? The first relevant question in this chapter is whether the RT effects of perceptually related pairs are due to timing constraints in the experiment or if they are indeed due to the perceptual or linguistic relationships between word pairs.

Subjects were predicted to be even more influenced to rely on quick linguistic representations, if the instructions asked them to make responses as quickly as possible, and would be more likely to rely on perceptual representations if they were asked to respond as accurately as possible. Importantly, responses that take longer would then be more likely to encourage the use of perceptual representations, which is consistent with prior research. Therefore, the prediction was that if participants were under a speed time constraint and instructed to be as fast as possible, then they would rely on linguistic information, such as statistical linguistic frequencies but if the participants were given more time or instructed to be more accurate they would access perceptual representations.

Method

The experiment was a 2x2 design where response speed (fast or slow) and instructions (accuracy-focus or speed-focus) varied.

Participants

Ninety-four native English speakers in the United States were recruited through Mechanical Turk (Mean Age = 34.68, *SD* = 12.12). Forty-five participants were randomly assigned to the fast response condition, 49 to the slow response condition, 43 (25 fast and 18 slow) to the accuracy-focus response condition, and 51 (20 fast and 31 slow) to the speed-focus response condition.

Table 1

Related and unrelated iconic word pairs and their LSA cosine values.

Related Word One	Related Word Two	LSA	Unrelated Word One	Unrelated Word Two	LSA
airplane	runway	0.77	belt	shoe	0.23
antenna	radio	0.74	billboard	highway	0.09
attic	basement	0.55	boat	trailer	0.06
car	road	0.43	bouquet	vase	0.28
ceiling	floor	0.72	branch	root	0.23
fender	tire	0.45	bridge	river	0.28
flame	candle	0.59	charcoal	grill	0.19
hat	scarf	0.47	cork	bottle	0.3
head	foot	0.46	cup	saucer	0.4
jockey	horse	0.43	faucet	drain	0.38
kite	string	0.45	flower	stem	0.27
knee	ankle	0.71	fountain	pool	0.31
lid	box	0.57	glass	coaster	0.07
mailbox	post	0.44	handle	bucket	0.2
mane	hoof	0.51	headlight	bumper	0.41
monitor	keyboard	0.51	hiker	trail	0.04
moustache	beard	0.61	hood	engine	0.23
nose	mouth	0.56	lamp	table	0.2
pan	stove	0.51	lighthouse	beach	0.37
pedestrian	sidewalk	0.45	mantle	fireplace	0.06
rocket	launchpad	0.68	penthouse	lobby	0.09
seat	pedal	0.44	pitcher	mound	0.16
smoke	chimney	0.51	plant	pot	0.18
steeple	church	0.52	sheet	mattress	0.12
stirrup	saddle	0.54	sky	ground	0.34
sweater	pants	0.61	sprinkler	lawn	0.11
track	runner	0.63	stoplight	street	0.01
train	railroad	0.66	tractor	field	0.2

Materials

The experiment consisted of 56 pairs of words that shared an iconic relationship (e.g., *cup* — *saucer*) where one item is found either above or below another (see Table 1). These word pairs were extracted from prior research (Louwerse & Jeuniaux, 2010). To reduce the likelihood of participants developing expectations about the experiment, 56 filler items consisted of word pairs without an iconic relation, with half of the pairs having a high semantic association and half having a low semantic association as determined by latent semantic analysis (LSA) (Landauer, McNamara, Dennis, & Kintsch, 2007). LSA is a computational linguistic technique that allows for estimating the relationship between words in a corpus while ignoring word order. Each participant saw half of the critical items in their iconic ordering (e.g., *cup* — *saucer*), and the other half in a reverse iconic order (e.g., *saucer* — *cup*), this was counterbalanced throughout the experiment.

Procedure

As in prior studies, participants were asked to judge the semantic relatedness of word pairs presented on a computer screen. Words were presented one above another in a vertical configuration, with the first word appearing at the top of the screen, and the second at the bottom. Upon presentation of a word pair, participants indicated whether the pair was related in meaning by pressing designated counterbalanced yes or no keys. All word pairs were randomly ordered for each participant to negate any order effects and each trial was separated by a '+' fixation symbol.

Subjects were also instructed to either respond as quickly as possible to the words they saw, or respond as accurately as possible to the words they saw. In the fast response condition, participants were allotted 1000 ms to respond to the stimuli before a message reading 'TOO SLOW' would appear in the center of the screen. In the slow response condition, participants were allotted 2500 ms to respond to the stimuli before a message reading 'TOO SLOW' would appear. In the accuracy-focus condition, participants were asked to try to be as accurate as possible in their responses, whereas in the speed-focus condition, participants were asked to try to be as quick as possible in their responses. Subjects were asked to describe the directions in a few sentences before beginning the task to ensure understanding. They were also asked to write a few sentences describing what they thought the purpose of the experiment was after they completed the session. There were no participants who misunderstood the directions, nor did any participants guess the true purpose of the experiment.

Results and Discussion

Four participants were removed from the analysis because more than 30% of their responses were incorrect, as measured by incorrect responses to lexical items. The linguistic and perceptual factors were operationalized as in Louwerse and Jeuniaux (2010), with the linguistic factor being calculated as the log

Table 2

Mean and SD values of RTs for accuracy, speed, fast, and slow conditions.

	Accuracy	Speed
Fast	*M* = 806.27 *SD* = 252.22	*M* = 787.12 *SD* = 258.81
Slow	*M* = 1328.47 *SD* = 473.27	*M* = 1247.89 *SD* = 454.96

frequency of each word pair, in both orders. The order frequency of all word pairs within 3–5 word grams was obtained using the large Web 1T 5-gram corpus (Brants & Franz, 2006). The perceptual factor was operationalized as an iconicity rating of the stimulus pairs whereby a set of different participants from the University of Memphis were asked to estimate the likelihood that the word pairs appeared above one another in in the real world. Ratings were made on a scale of 1–6, with 1 being extremely unlikely and 6 being extremely likely.

All analyses were mixed models that specified participants and items as random factors and RT as the dependent variable. To ensure participants correctly performed their task speed was tested for and a main effect was found, $F(1, 9378) = 38.43$, $p < .01$, $\eta^2 = .004$ with faster RTs in the fast time constraint. A main effect of task was also found, $F(1, 9378) = 20.95$, $p < .01$, with faster RTs in condition where participants were asked to respond as quickly as possible. Furthermore, there was a significant interaction between time constraint speed and task, $F(1, 22.45) = 26.96$, $p < .01$, $\eta^2 = .55$ (see Table 2).

These findings suggest that participants were faster to respond to word pairs during a semantic judgment task when a 1000 ms time constraint was imposed than when they were allotted 2500 ms to respond. In addition, participants were faster to respond to word pairs when they were asked to focus on responding quickly than when they were asked to focus on responding accurately.

To determine if linguistic and/or perceptual factors impacted processing during a semantic judgment task, for all correct critical trials, a linear mixed effect model with word frequency and perceptual ratings as fixed factors and participant and item as random factors was run. Word frequency best explained resulting RTs, $F(1, 2349) = 26.96$, $p < .01$, $\eta^2 = .01$, although the perceptual factor also contributed (albeit not significantly), $F(1, 2349) = 2.93$, $p = .08$, η^2

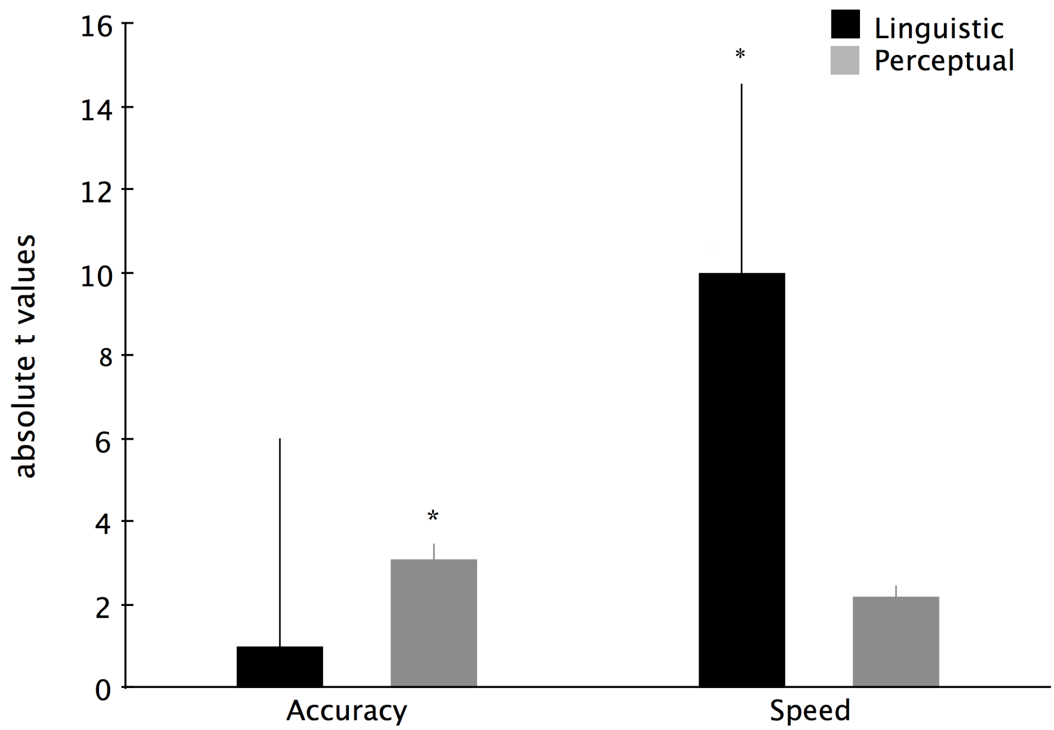


Figure 1. Strength of the effect in absolute t values of perceptual and linguistic factors for accuracy and speed conditions. * denotes $p < .05$

= .001. These findings suggest that the linguistic factor accounts for processing during a semantic judgment task. However, past research found that both linguistic and perceptual factors explained RTs (Louwerse & Jeuniaux, 2010). Perhaps the perceptual factor here failed to reach significance for all participants due to the fact that many of the participants were under time constraints.

In fact, when participants were asked to focus on accuracy, word frequency did not explain RTs $F(1, 1177) = 1.75$, $p = .19$, $\eta^2 = .001$, while the perceptual factor did, $F(1, 1177) = 8.08$, $p < .01$, $\eta^2 = .001$ (see Figure 1). This is in line with the idea that perceptual simulations are more relevant later during processing. However, when participants were asked to focus on speed, word frequency explained RTs, $F(1, 1169) = 97.42$, $p < .01$, $\eta^2 = .08$, but the perceptual factor did not, $F(1, 1169) = 2.36$, $p = .12$, $\eta^2 = .002$. These results suggest that linguistic factors might play a more important role early, and perceptual information becomes important later in processing.

To determine if this is the case, a linear mixed effect regression was run where word frequency and a perceptual factor were fixed factors and participants and items were random factors for analyses of both the short and long periods of time. When participants were only given a short time period to respond, word frequency accounted for RTs, $F(1, 1125) = 7.01$, $p < .01$, $\eta^2 = .006$, whereas the perceptual factor remained irrelevant $F(1, 1125) = 0.31$, $p = .57$, $\eta^2 < .001$. In contrast, when participants were given a longer time period to respond, both factors explained RTs, with word frequency $F(1, 1221) = 5.92$, $p = .02$, $\eta^2 = .005$, and the perceptual ratings $F(1, 1221) = 3.67$, $p = .05$, $\eta^2 = .003$, both reaching significance (see Figure 2). These results are in line with prior research that suggest that when participants are given enough time to respond to word pairs in a semantic judgment task, linguistic factors and perceptual factors are relevant for processing, with linguistic representations preceding perceptual representations (as the linguistic factor was significant for the short time period, whereas the perceptual factor only became significant for the longer time period).

So it might be the case that RT effects are simply influenced by the amount of time a participant used to respond to a word pair, as time constraints

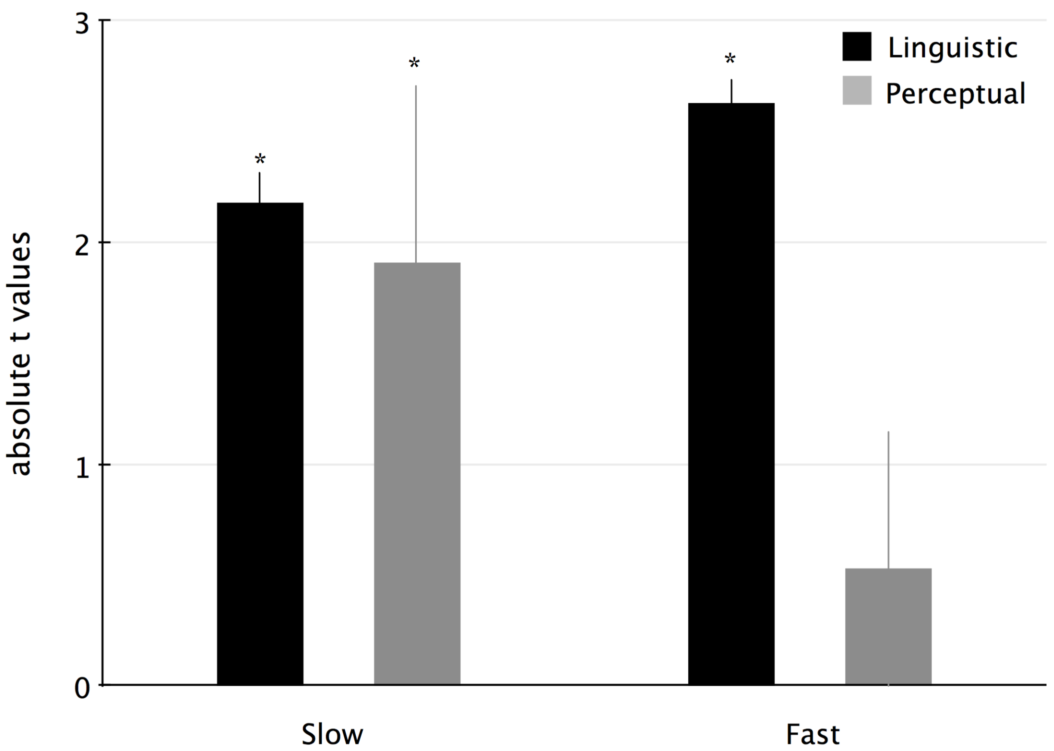


Figure 2. Strength of the effect absolute t values of perceptual and linguistic factors for accuracy and speed conditions. * denotes $p < .05$

influenced a participant's reliance on perceptual or linguistic factors. Put differently, it seems as if we are using perceptual simulations when we make slower decisions and likewise for linguistic representations and fast decisions. In this chapter, RT effects of perceptually related pairs were indeed influenced by the timing constraints in the experiment and not simply from the perceptual or linguistic relationships between word pairs. Specifically, participants were more likely to utilize perceptual information when they had more time to process word pairs on a screen. Conversely, participants relied more on linguistic information (and less on perceptual information) when they had more stringent timing constraints. In comparison, if it was instead the case that embodied (or linguistic) effects are found only because of an embodied (or linguistic) relationship between word pairs, timing constraints are not expected to impact the effect of either perceptual or linguistic representations. Importantly, both factors are relevant throughout the time course of processing word pairs, with the linguistic factor being relatively *more* important early during processing and perceptual information being relatively *more* important later. In fact, the linguistic factor significantly explains RTs in both the slow and fast conditions. Since the linguistic representations precede perceptual simulation during processing (Louwerse & Hutchinson, 2012), it is logical that the linguistic factor would still remain relevant in later processing, with the perceptual factor becoming relatively more relevant.

Experiment 2: Space

Although it is evident from Experiment 1 that linguistic representations are more prominent early in processing of word pairs, and perceptual representations are more prominent later, it remains difficult to offer a linguistic explanation for results when words are presented in isolation. In Experiment 1, words were presented in pairs but embodied cognition studies have demonstrated that when individual words found in high physical locations (e.g., *bird*) are positioned at the top of a screen they are processed faster than when they are positioned at the bottom of the screen. The reverse effect is obtained for words found in low physical locations (e.g., *fish*). This concept-location facilitation effect has been argued to demonstrate that cognitive processing is fundamentally perceptual in nature. Although task instructions might influence the speeded responses, the frequency of *butterfly — sky* is only able to account for faster RTs for congruent word categories and tasks while still leaving mental simulations to offer the only explanation for the facilitative effect of the congruency of the presentation location and the perceptual features of the word (as unigram word frequency cannot account for these RTs). However, questions can be raised with regard to the absolute or relative location of these concept-location words. Perhaps linguistic information might play a role explaining these concept-location effects for isolated words after all.

Even though words are presented in isolation on the screen (i.e., one word is presented at a time), it is possible that decisions might be made relative to the other words presented in the other trials of the experiment. Such an explanation would suggest that instead of making judgments relative to the congruency between the concept and the absolute position of the word on the screen (i.e., top of the screen or the bottom of the screen), participants are making judgments relative to the other words in the experiment. That is, participants might show a concept-location facilitation effect not because the words are presented on the top and bottom of the screen, but rather because words are asynchronously presented relatively above and below one another throughout the duration of the experiment. Although this may seem to be a straw man argument, a number of embodied cognition researchers indeed claim that it is important that the facilitative effect between the spatial location of the word is influenced by its absolute location, "without relying on the relative position of [the] word" (Šetić and Domijan, 2007, p. 300).

To explore this possibility, participants were presented with isolated words at either the top or bottom (to replicate the original results), top or center, or center or bottom of the screen. According to an embodied cognition account, if responses are faster because word meaning and world location are congruent, the same high and low words, presented in the center of the screen should show no concept-location facilitation effect because the presentation location is not congruent with the physical and spatial properties of the simulated word. In other words, when *butterfly* is presented in the center of the screen, processing should not be facilitated. Alternatively, if decisions are based on the relationship between one word relative to the other words in the experiment (as opposed to being relative to the presentation location of the word; a linguistic explanation), then high words presented in the center of the screen (concept-location mismatch) should still show a concept-location facilitation effect if low words are presented at the bottom of the screen. That is, when *butterfly* is presented in the center of the screen, processing will be facilitated if other words in the experiment are ‘below’ a butterfly. Similarly, low words presented at the center of the screen would show a concept-location facilitation effect if high words are presented at the top of the screen.

In essence, if concept-location facilitation is found when words are presented in relative positions on the screen (i.e., above/below one another) as opposed to absolute positions on the screen (i.e., at the top/bottom of the screen), it might be the case that perceptual simulation (concept-location facilitation effect) is not entirely accounting for RTs but rather, participants are making decisions about words presented in isolation by comparing those words to the group of words included in the experiment. The following experiment explores whether semantic judgments are made with respect to an absolute location on the screen (embodied explanation) or with respect to a relative location in comparison to other words included in the experimental session (statistical linguistic explanation) in order to determine if spatial location of the stimuli, like temporal constraints, can also impact how likely participants are to rely on linguistic or embodied mental representations. In a response time experiment, participants were presented with physical-location words from existing studies at the top or bottom, top or center, and center or bottom of the screen.

Participants

Eighty-seven undergraduate native English speakers at the University of Memphis participated for extra credit in a Psychology course. Participants were randomly assigned to each of the three conditions (words presented at either a) the top of the screen and the center of the screen, b) the center of the screen and the bottom of the screen, or c) the top of the screen and the bottom of the screen).

Materials

The experiment consisted of 48 living animal words that could be found in a low spatial location, (such as the ground or ocean, n=24) or found in the sky (a high spatial location, n=24). The remaining 48 words consisted of non-living objects that could also be found in either high (n=24) or low (n=24) physical locations. Words were extracted from both Pecher et al. (2010) and Šetić and Domijan (2007).

Procedure

The procedure was almost identical to Pecher et al. (2010) and Šetić and Domijan (2007). Participants were asked if words presented on a 1280x1024 computer screen were either living or nonliving. This task has the advantage that it does not bias participants to consciously judge the physical location of a word. The center of the screen was positioned at eye level. Similar to Pecher et al. (2010) and Šetić and Domijan (2007), each trial began with the presentation of three fixation crosses appearing on the screen for 300 ms. Fixation crosses were presented either at the top, center, or bottom of the screen, depending on where the proceeding word would appear on the screen. This occurred in order to notify participants where the next word would appear.

Words were presented at either the top and the center of the screen, the center and bottom of the screen, or — as in the original Šetić and Domijan (2007) study the top and bottom of the screen, depending upon the between participants condition. Upon presentation of a word, participants indicated whether the word was living or not living by pressing designed counterbalanced keys on the keyboard (*f* and *j* keys). All words were seen once and were counterbalanced for each participant where half the high spatial location words were presented in the upper position (relative to the other presentation location, i.e., top relative to center/bottom or center relative to bottom) and half in the lower position (i.e., bottom relative to center/top or center relative to top), likewise for the low spatial location words.

If responses were slower than 2,500 ms a message reading ‘TOO SLOW’ would appear. Participants were asked to try to be as quick and as accurate as possible in their responses. The next trial began immediately after the participant's response or after the feedback message.

Results and Discussion

Eleven participants were removed from the analysis because >40% of their answers were incorrect. All remaining participants were split evenly between conditions. In all analyses, the parameters found in Pecher et al. (2010) were used for outlier identification and removal. Outliers were identified as those correct responses greater than three standard deviations from the mean per participant per item. Outlier removal (as described above) resulted in a loss of 2.8% of the data. All error trials were removed, resulting in a loss of an additional 8.7% of the data.

A mixed-effect regression analysis was conducted on RTs with match/mismatch (match or mismatch between word category (low or high spatial location word) and relative presentation location (relatively high location of top or center or relatively low location of center or bottom) as a fixed factor and participants and items as random factors. The model was fitted using the restricted maximum likelihood estimation (REML) for the continuous variable (RT). F-test denominator degrees of freedom were estimated using the Kenward-Roger's degrees of freedom adjustment to reduce the chances of Type I error (Littell, Stroup, & Freund, 2002).

In addition to the location presentation manipulation, the source of the RT differences in this task was also investigated, linguistic or embodied. An embodied account would predict a concept-location facilitation effect, whereas a linguistic account would suggest these same effects are driven by language statistics. To further explore if participants were relying on language statistics, analyses using word frequency as a fixed factor were used to determine if a possible additional explanation for any concept-location facilitation effects might exist. The word frequency factor was calculated as the log frequency of each word being presented obtained using the Web 1T 5-gram corpus (Brants & Franz, 2006).

Unlike Šetić and Domijan (2007), no significant concept-location facilitation effect was found for words appearing at the top of the screen, $F(1, 2330) = 1.46$, $p = .23$, $\eta^2 < .001$, at the center of the screen, $F(1, 1599) = .10$, $p = .75$, $\eta^2 < .001$, nor at the bottom of the screen, $F(1, 2395) = 1.76$, $p = .19$, $\eta^2 < .001$. Just as in Pecher et al., (2010) these findings also fail to replicate the concept-location facilitation effect found in Šetić and Domijan (2007). In fact, there was no interaction between location and word category for any of the three word presentation locations and experimental conditions. Pecher et al. (2010) offered the explanation that the concept location facilitation effect is not well understood, with some factors causing facilitation and others causing interference. The linguistic frequency factor did not explain the results either, with no significant main effects for words appearing at the top of the screen,

$F(1, 2330) = .0001$, $p = .99$, $\eta^2 < .001$, the center of the screen, $F(1, 1599) = .19$, $p = .66$, $\eta^2 < .001$, nor the bottom of the screen, $F(1, 2395) = .11$, $p = .74$, $\eta^2 < .001$. These current results seem to support neither an embodied cognition account (as there was no concept-location facilitation for the top-bottom condition) nor an alternative linguistic account (as there was no concept-location facilitation for either condition including the center location nor was linguistic frequency significant). In the absence of a replication in both the current study and in Pecher et al. (2010), perhaps the effects reported in Šetić and Domijan (2007) might be attributed to linguistic differences in the Hungarian stimuli. Alternatively, such concept-location facilitation effects might simply be relevant for certain groups of words and not others.

To further explore the results of the current experiment, and the possibility that words are processed relative to the words around them, results were analyzed for animate versus inanimate words. Words that were inanimate again showed no interactions for words appearing at the top of the screen, $F(1, 1172) = .003$, $p = .96$, $\eta^2 < .001$ (see Figure 3), the center of the screen, $F(1, 787) = .07$, $p = .80$, $\eta^2 < .001$ (see Figure 4), or the bottom of the screen, $F(1, 1072) = .92$, $p = .34$, $\eta^2 < .001$ (see Figure 5). Linguistic frequency was also not significant for words appearing at the top of the screen, $F(1, 1172) = 1.53$, $p = .22$, $\eta^2 = .001$, the center of the screen, $F(1, 787) = .62$, $p = .43$, $\eta^2 < .001$, nor the bottom of the screen, $F(1, 1072) = .002$, $p = .96$, $\eta^2 < .001$.

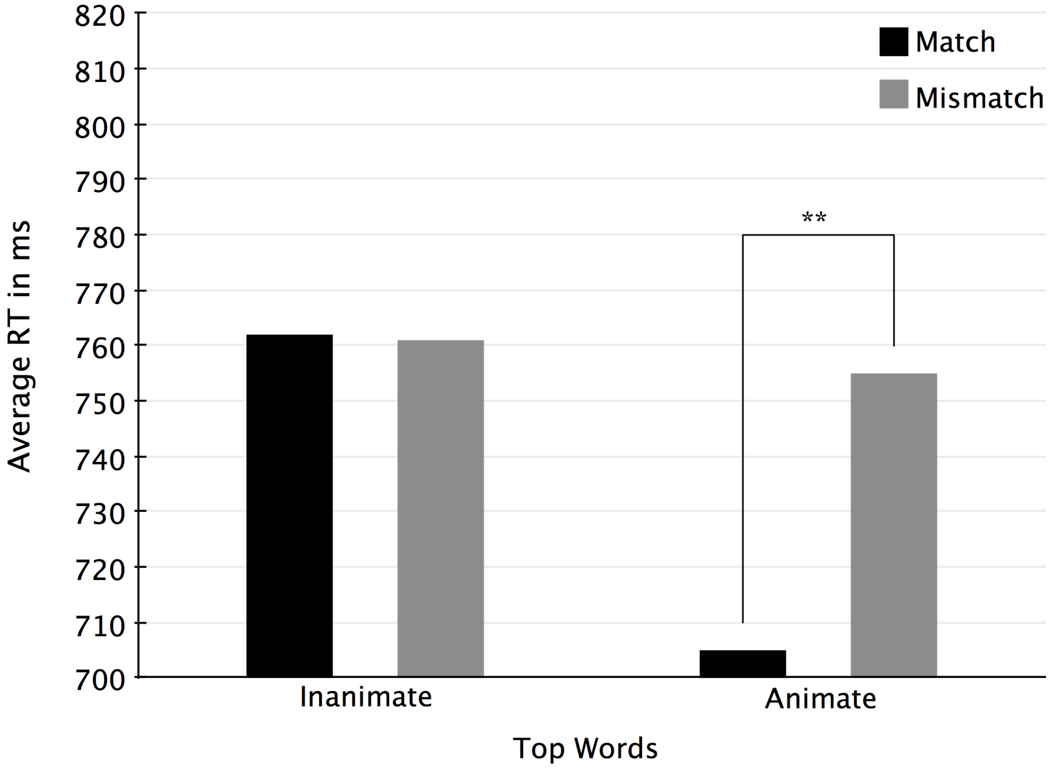


Figure 3. Average RTs in ms for the words appearing at the top of the screen.

However, words that were animate did show significant interactions. Words appearing in any given location (top, center, and bottom) were processed faster when that location was relatively the same as the word category. ‘Up words’ presented in the center were processed faster in the center-bottom condition, whereas ‘down words’ presented in the center were processed faster in the top-center condition, $F(1, 789) = 6.10$, $p < .02$, $\eta^2 = .008$. Figure 4 shows RTs for matched and mismatched up and down words presented in the center of the screen, showing that words with a concept-location match are processed faster than words with a concept-location mismatch. Similarly, ‘up words’ presented in the top of the screen were processed faster in both the top-bottom and top-center conditions, $F\ (1, 1134) = 6.80$, $p < .01$, $\eta^2 = .006$ (see Figure 3).

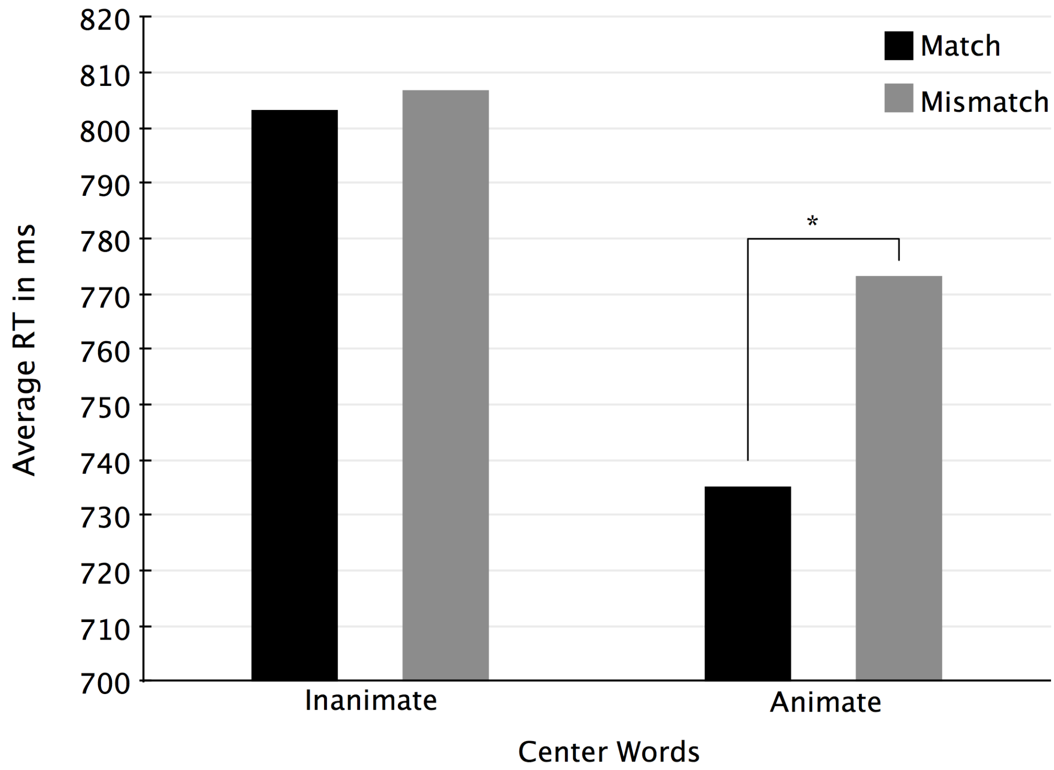


Figure 4. Average RTs in ms for the words appearing at the center of the screen.

Finally, 'down words' presented in the bottom of the screen were processed faster in both the top-bottom and center-bottom conditions, $F(1, 1067) = 10.97$, $p = .001$, $\eta^2 = .01$ (see Figure 5).

In addition, to further explore the impact of linguistic frequency also significantly explained RTs to words presented at the bottom of the screen, $F(1, 1067) = 5.08$, $p = .02$, $\eta^2 = .004$, but only marginally for words presented in the center of the screen, $F(1, 789) = 3.22$, $p = .07$, $\eta^2 = .004$, with no effects for words presented at the top of the screen, $F(1, 1134) = 2.58$, $p = .10$, $\eta^2 = .002$. These findings seem to be consistent with the idea that decisions are based on the relationship between one word relative to the other words in the experiment, as 'up words' presented relatively above 'down words' still showed a concept-

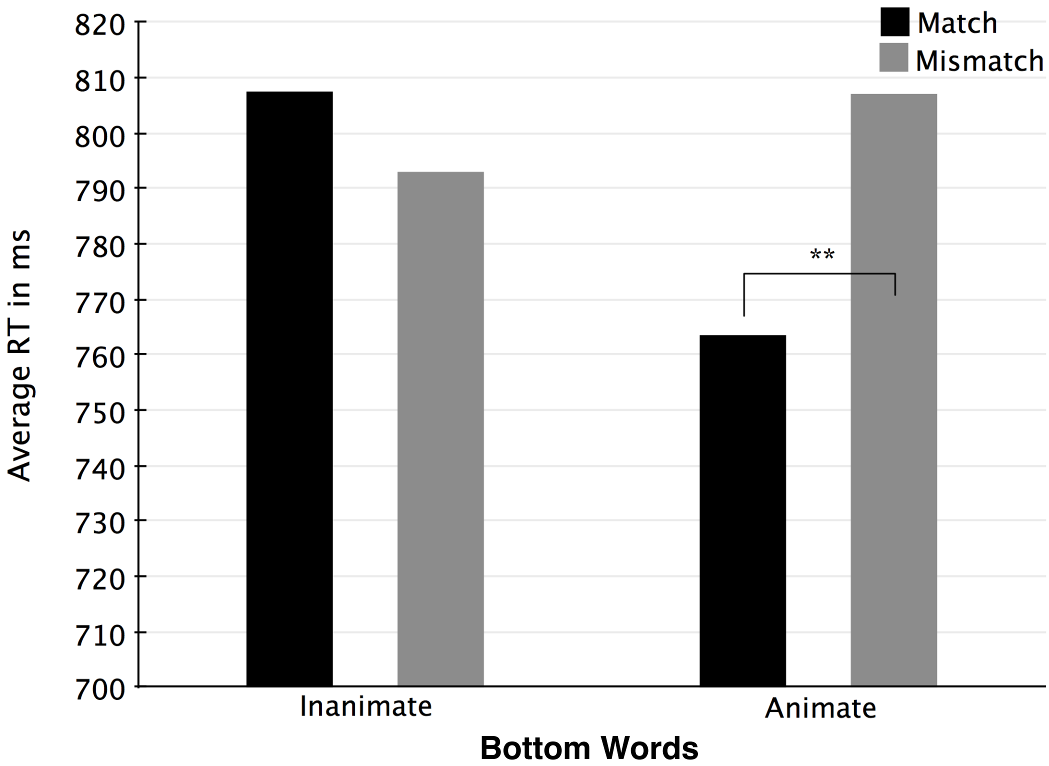


Figure 5. Average RTs in ms for the words appearing at the bottom of the screen.

location facilitation effect despite these words being presented in the center of the screen.

In addition, in all conditions, words appearing relatively below other words (M= 767.45, SD=267.40) were processed significantly slower than words appearing relatively above other words (M= 889.36, SD=421.41), $t(4926) = 15.36$, p <.001. That is, regardless of the absolute location of the word on the screen, where ever the bottom position was (i.e., center of the screen or bottom of the screen), words presented in that location were processed slower than the same words presented in a relatively higher location. Consider the case of the center presentation location: when words were presented in either the center of the screen or the bottom of the screen, words took longer to process at the bottom and less time to process at the center. However, when those same words were presented in the center or the top, they took longer to process in the center and less time to process at the top. This means that the same words presented in the same location are processed faster or slower simply due to whether other words are appearing above or below them. This at least suggests that comparisons between high and low positions are biased given that the center represents both the relative top and bottom in different conditions.

Finally, to explore whether participants indeed made comparative judgments for words, bigram frequencies were assessed and were able to account for the response times of center words. As in previous studies (Louwerse, 2008) bigram linguistic frequencies were operationalized as the log frequency of a-b (e.g., *owl-lizard*) or b-a (e.g., *lizard-owl*) order of word pairs. Because words were presented individually on the screen, pairs were determined by the randomized presentation order. The bigram frequency of each pair was assigned to the second word in the randomly presented pair. The order frequency of all word pairs within 3-5 word grams was obtained using the large *Web 1T 5-gram* corpus (Brants & Franz, 2006). A mixed-effect regression analysis was conducted on RTs to center words with the bigram frequency as a fixed factor and participants and items as random factors (Baayen, Davidson, & Bates, 2008). Bigram frequency was a significant predictor of RTs for center words only, $F(1,906) = 3.99$, $p = .05$, $\eta^2 = .004$. This was true for all center words regardless of experimental condition, implying that participants consider past trials while making judgments about the current word in question, and implying that a linguistic frequencies explain RTs during a concept-location facilitation task.

The absolute location of a word on a screen does not seem to impact the concept-location facilitation effect, but rather the relative location appears to be what is important. This finding suggests that decisions are based on the relationship between one word relative to the other words in the experiment, not only based on the relationship between one word and the embodied physical and spatial properties of that simulated word. In addition, across all three conditions, there was a main effect of location, such that words presented below other words were processed slower. This finding suggested that participants made judgments relative to other words, not only relative to their location on the screen. To further determine whether participants made comparative judgments between words presented asynchronously over the duration of an experiment bigram frequencies were shown to predict participant RTs. These findings together indicate that it might be the case that participants are making decisions about words presented in isolation by comparing those words to the group of words included in the experiment, suggesting that findings that are easily attributed to embodied cognition (Pecher et al., 2010; Šetić & Domijan, 2007) can also be attributed to language statistics.

Experiment 3: Relating the unrelated

The central research question at the start of this dissertation concerned the extent to which perceptual or linguistic representations are relied upon during processing. In the previous two experiments, I demonstrated that this is indeed the case, with time and space influencing when and how participants rely more or less on linguistic versus perceptual representations during processing. This finding is crucial for Experiment 3 in which, relying on the assumption from Experiment 1, that linguistic processing is more important in the early stages of language processing, and perceptual processing is more important later, I demonstrated that both linguistic and perceptual representations, although intertwined are still independent processes. I began by examining why these two processes might be mistaken for being wholly dependent upon one another, then moved on to seeing how these processes are indeed independent while still remaining highly related.

Such a strong relationship between linguistic and perceptual representations has lead embodied cognition theorists to argue that perhaps linguistic representations are not independent at all, but rather artifacts of the perceptual system. Despite evidence pointing in favor of complimentary but also independent linguistic processing, the question of whether linguistic representations simply emerge from the use of perceptual representations is still unanswered. Logically, when processing the word *strawberry* both related words like *pie, shortcake, blonde,* and perceptual simulations of the color, size, and taste of a strawberry are related. It is difficult to disentangle the boundaries between linguistic representations and perceptual representations, for this very fact, because they refer to each other.

In fact, for many concept words used in embodied cognition experiments, the frequency pattern matches the perceptual relation (Louwerse, 2008), suggesting that perceptual and linguistic systems do indeed activate simultaneously to give meaning to words, with linguistic representations being more prominent during early processing and perceptual representations reaching dominance later (Hutchinson, Tillman, & Louwerse, 2014; Louwerse & Hutchinson, 2012; Louwerse & Jeuniaux, 2010). However, both types of relationships are active throughout language processing, only their importance modulates over time. The following experiment is designed to show that the linguistic system acts independently of the perceptual system.

In order to determine if both linguistic and perceptual representations are activated when generating relationships between word pairs, a RT experiment was designed whereby participants were asked either to determine whether linguistically and/or perceptually similar or dissimilar word pairs were semantically related. Following the aforementioned research and prior work, I relied on the assumption that linguistic processing would occur relatively more early during language processing and perceptual processing would occur relatively later during processing.

It was hypothesized that if linguistic features dominate early processing and perceptual features become salient later, as shown in Experiment 1, then it follows to expect participants to respond to linguistically related pairs with lower RTs than to pairs lacking such a relationship because if word pairs only share a linguistic relation, then a linguistic representation should suffice to garner meaning. It was therefore predicted that perceptually and linguistically similar word pairs should be just as easy to process as pairs that are only linguistically similar.

Further, it was predicted that any word pair that shares a linguistic relationship should be processed with lower RTs than word pairs that share only perceptual similarities. In other words, if word pairs only share a perceptual relation, then a linguistic representation would not be sufficient in order to determine the relationship between the word pair because such a representation might lack modality-specific perceptual information. Instead, a complete perceptual representation would be necessary to compare the perceptual features of both words and in turn determine their perceptual relationship.

An interesting question arises in how participants will tend to relate dissimilar pairs, i.e., when attempting to associate two unrelated items, will participants tend to rely upon the linguistic features or the perceptual features of those items to generate a relation between them? For seemingly unrelated word pairs, a linguistic representation might be expected to be insufficient to generate a relation between the pair, although the linguistic context still remains important. Instead consideration of all features of the word would be necessary to extrapolate a relationship for the pair. If associating dissimilar pairs does indeed take longer than processing pairs that are related, then one might predict that, in line with the notion that perceptual features are more salient in later processing, that participants generate a full perceptual representation when trying to relate dissimilar words. Due to this fact, one would expect perceptual factors to better predict RTs for unrelated pairs, and for those RTs to be longer than those of semantically related pairs. If words as unrelated as *strawberry* and *pony* are processed in a similar fashion to words with a semantic or perceptual relationship, then are perceptual representations fundamental for language processing, or rather simply beneficial for language understanding?

If words as unrelated as *strawberry* and *pony* are processed in a similar fashion to words with a linguistic or perceptual relationship, then are perceptual representations fundamental for language processing, or rather simply beneficial for language understanding? Note that all of the experiments prior hinge on the symbolic factor being mostly independent from the embodied factor. This chapter thereby attempts to disentangle whether linguistic and perceptual representation are dependent on one another by demonstrating that in some cases (i.e., with highly related word pairs), linguistic representations might just be good enough for meaning and in other cases (i.e., with highly unrelated word pairs), both linguistic and perceptual representations are necessary for comprehension.

In the following experiment, participants were asked either to determine whether the words presented were semantically related. Because linguistic features dominate early processing and perceptual features become salient later (Louwerse & Connell; 2011), participants would be expected to respond to semantically related pairs with lower RTs than to pairs lacking such a relationship. If word pairs only share a semantic relation, then a linguistic representation would suffice to gather meaning and assess the relationship between the words. Perceptually and semantically similar word pairs were predicted to be just as easy to process as pairs that were only semantically similar. Further, any pair with a linguistic relationship is predicted to be processed with lower RTs than pairs with only perceptual relationships. If the word pairs only shared a perceptual relation, then a linguistic representation would not be sufficient in order to determine the relationship between the word pair because such a representation might lack modality-specific perceptual information. Instead, a complete perceptual representation would be necessary to compare the perceptual features of both words and in turn determine their perceptual relationship. Furthermore, linguistic factors are expected to best predict performance for semantically related pairs, whereas both linguistic factors and perceptual would be expected to best predict performance on related pairs. Finally, following the hypothesis that a full perceptual simulation is required to generate a relationship between unrelated pairs, perceptual factors would be expected to best predict performance on unrelated pairs.

Participants

Thirty-seven participants were recruited from mTurk online. All participants were native English speakers.

Materials

Each experiment consisted of 62 trials, each including a pair of words. Word pairs were presented in a random order, while the word order making up each pair was counterbalanced between participants. Fifteen word pairs were linguistically and perceptually similar, 15 word pairs were only linguistically similar, 15 word pairs were only perceptually similar, and 15 word pairs were linguistically and perceptually dissimilar. Some word pairs were extracted from previous research (Louwerse & Jeuniaux, 2010) and some word pairs were newly created. Linguistically related word pairs were determined by latent semantic analysis (LSA; Landauer, McNamara, Dennis, & Kintsch, 2007) (high semantic association: cos = .42; low semantic relation: cos = .07). In addition, the linguistic relatedness between word pairs was determined using the log frequency of all word pairs within 3-5 word grams was obtained using the large *Web 1T 5-gram* corpus (Brants & Franz, 2006) (high semantic association: log frequency = 11.83; low semantic relation: log frequency = 5.19). All words had concreteness and imageability scores greater than 400 (as determined by the MRC Psycholinguistic Database, a dictionary containing 150837 words with linguistic, psycholinguistic, and psychological measures; Coltheart, 1981) and were between 2-11 letters and all words pairs were rated on the number of shared perceptual features (e.g., shape, color, texture, function, size, taste, etc.). Perceptually related word pairs all had on average 2.58 (*SD* = 1.34) overlapping perceptual features and perceptually unrelated word pairs had at most one overlapping perceptual feature (*M* = .12, *SD* = .33).

Procedure

Word pairs were presented next to one another in the center of a computer screen. Participants were instructed to indicate a ‘yes’ or ‘no’ answer to the question “Are the words related in meaning?”. They were given examples of words that were and were not related in meaning, before starting the experiment. Participants were asked to try as best they could to come up with any type of relation between the two words and then pressed a designated yes or no button on the keyboard. Once a participant responded, the fixation symbol ‘+’ would appear center-screen for 1000 ms, followed by the next trial. Pairs within an experimental session were randomly presented in order to negate order effects. To ensure understanding, a session of six practice trials preceded the experimental session.

Results and Discussion

Three participants were removed from the analysis because they explicitly stated that they were relating words based on orthographical and lexical features. This resulted in a total of 34 participants where all remaining participants were split evenly between conditions. Because participants were asked to generate relationships between unrelated words, there were no incorrect answers for unrelated words, as the task was subjective. In order to filter participants who may not have been properly performing the task, RT outliers were examined. RT outliers were identified as those correct responses to related words (or any response to unrelated words) greater than 2.5 standard deviations from the mean per participant per item. Outlier removal resulted in a loss of 5.8% of the data.

First, mixed models were conducted to compare RTs, participant and item were specified as random factors. RTs significantly varied between word pairs, $F(1, 2031) = 222.606$, $p < .001$, , $\eta^2 = .10$ (see Table 3). To explore whether RTs were faster when processing word pairs sharing a linguistic relationship, or a perceptual relationship, a linear mixed model was run whereby perceptual relatedness and linguistic relatedness were treated as categorical fixed factors, and subject and item were treated as random factors. Whether or not the word

Table 3

Mean RTs and SDs to various stimuli pairs

Type of Relationship	RT
Unrelated	$M = 2178$ ($SD = 1405$)
Perceptual	$M = 1861$ ($SD = 1054$)
Semantic	$M = 1202$ ($SD = 599$)
Semantic and Perceptual	$M = 1170$ ($SD = 594$)

pair was related linguistically significantly impacted RTs, $F(1, 2031) = 252.14$, $p < .001$, as did whether or not the word pair was related perceptually, $F(1, 2031) = 24.17$, $p < .001$, $\eta^2 = .018$. Furthermore, there was a significant interaction between the two, $F(1, 2031) = 16.10$, $p < .001$, $\eta^2 = .008$.

These results demonstrate that word pairs sharing any type of relationship were processed faster than those unrelated word pairs, $F(1, 2031) = 70.12$, $p < .001$, $\eta^2 = .033$. However, not only were related word pairs processed faster than unrelated word pairs, but those pairs that shared any type of linguistic relationship were processed quickly, regardless of their perceptual relation, whereas those pairs that only shared a perceptual relationship were processed

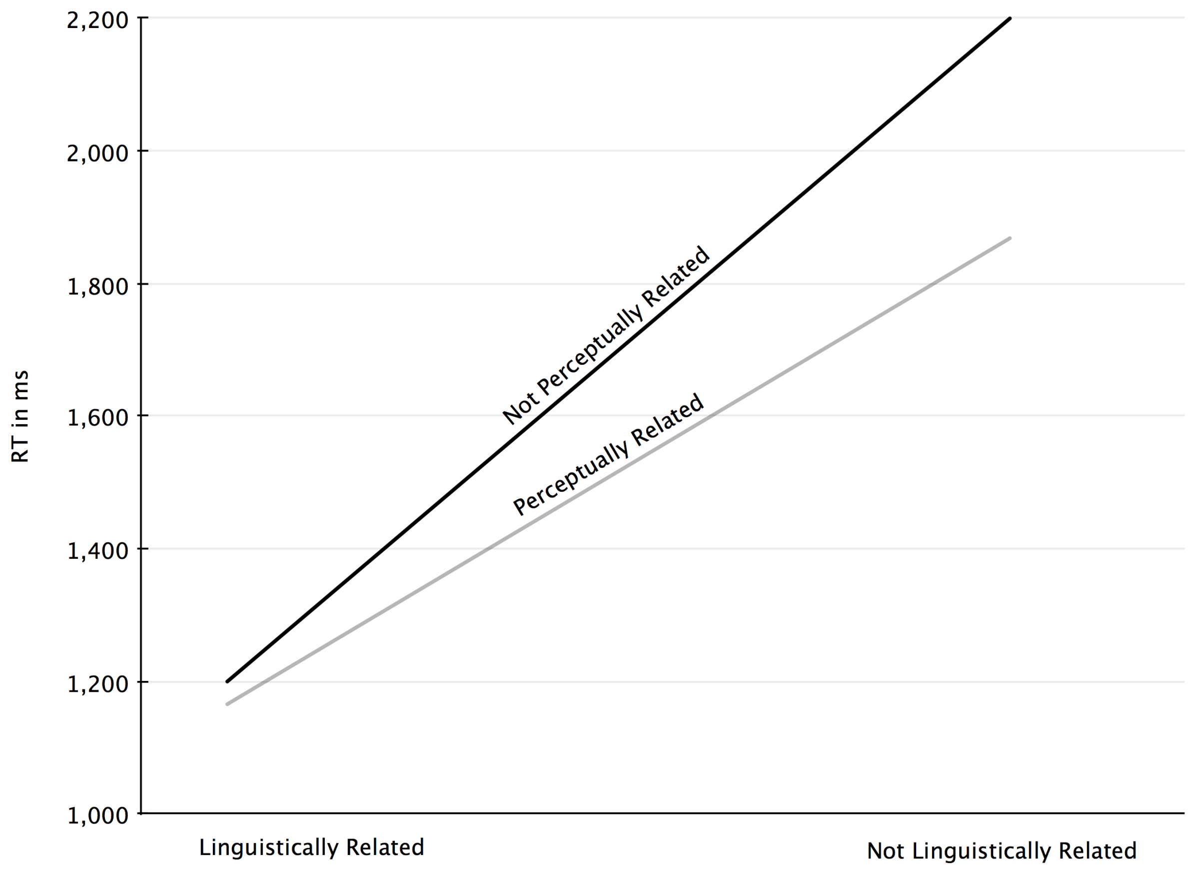


Figure 6. Linguistically and perceptually related and unrelated word pair RTs.

significantly slower than any linguistically related pairs, $F(1, 1510) = 164.87$, $p < .001$, $\eta^2 = .10$. This is in line with the prediction that participants would respond to linguistically related word pairs with lower RTs than to perceptually related word pairs, as linguistic processing precedes perceptual simulation and therefore linguistically related word pairs should be processed faster.

Furthermore, because linguistic relatedness is sufficient to determine the relation between word pairs, pairs that only shared a linguistic relationship were expectedly processed no faster than pairs that shared a linguistic and perceptual relationship, $F(1, 1045) = 0.588$, $p = .449$, $\eta^2 < .001$. This suggests a linguistic representation might be sufficient to generate a relationship between the perceptually and linguistically similar word pairs. In other words, a full perceptual simulation might not be necessary when making a judgment about the relatedness of word pairs that are both perceptually and linguistically related.

To further explore whether RTs were actually explained by linguistic or perceptual information, mixed models were run where linguistic frequency and perceptual ratings were entered as fixed factors, and subject and item were considered random factors. Perceptual ratings were operationalized as the number of perceptual features (e.g., shape, color, texture, function, size, taste, etc.) shared between word pairs. The linguistic frequency for word pairs was calculated as how frequently *word1* and *word2* appeared together in a five-word window in the large *Web 1T 5-gram* corpus (Brants & Franz, 2006). The correlation between frequencies and ratings was significant at $r(2130) = .54$, $p < .001$. Given that language encodes perceptual information, this is not surprising, but presents a problem for researchers trying to differentiate between the two factors.

As expected for pairs that were semantically related, linguistic frequency moderately predicted RTs, $F(1, 538) = 3.03$, $p = .087$, $\eta^2 = .006$, whereas perceptual ratings did not predict RTs, $F(1, 538) = 1.99$, $p = .18$, $\eta^2 = .004$. For pairs that were perceptually related, linguistic frequency predicted RTs, $F(1, 1510) = 38.81$, $p < .001$, $\eta^2 = .025$, as did perceptual ratings to a lesser degree, $F(1, 1510) = 5.68$, $p = .02$, $\eta^2 = .004$. But importantly, for unrelated pairs, linguistic frequency did not predict RTs, $F(1, 484) = .009$, $p = .92$, $\eta^2 < .001$. However perceptual explained much more variance, $F(1, 484) = 4.008$, $p = .06$, $\eta^2 = .008$. This suggests that participants generate a full perceptual representation when trying to relate dissimilar words, as perceptual relationships are independently utilized when generating a relationship while linguistic representations are not.

Conclusion

In the current chapter, the three objectives were to determine a) if time constraints on a semantic judgment task could influence how much a participant relied on linguistic and perceptual factors during processing, b) whether participants made judgments relative to other words on the screen or relative to their absolute location on the screen, thereby establishing if spatial location of stimuli can also impact how much participants rely on perceptual or linguistic representations and c) if linguistic and perceptual representations are independent of one another.

In Experiment 1, the Symbol Interdependency Hypothesis predicted that linguistic factors are important immediately during processing, preceding a deeper perceptual simulation system. Results from a RT study found exactly that: participants relied more on a linguistic factor during processing when participants given strict time constraints, or when they were told to focus on responding quickly. When given more time to respond, both linguistic and perceptual factors explained response times. These findings are in line with the findings from Louwerse and Jeuniaux (2010), Louwerse and Connell (2011), and Louwerse and Hutchinson (2012) that suggest the linguistic representations are more relevant early on, and that perceptual representations are more relevant as time progresses. In a second experiment, three presentation location conditions (top and center, bottom and center, or top and bottom) failed to replicate a concept-location facilitation effect as found in Šetić and Domijan (2007) for inanimate words. However, when considering animate words, words matched between the relative presentation location and word category resulted in faster RTs than words with a mismatch. This finding suggests that participants make judgments about individual words they see on the screen with respect to other words they see throughout the duration of an experiment.

Together, these experiments suggest that multiple factors can influence how reliant we might be on different kinds of mental representations under different conditions, while emphasizing that both linguistic and perceptual representations are utilized during processing.

Importantly, the previous findings that have given support to perceptual simulation (e.g., Barsalou, 1999; Glenberg, 1997; Pecher & Zwaan, 2005; Semin & Smith, 2008; Zwaan, 2004) are not invalidated through these findings. In those instances, perceptual simulation was more suited for the task and presumably temporal and spatial constraints were not studied factors. However, the results from the current study show that perceptual simulation does not always win the struggle for the most efficient type of processing. The results of the current studies show not only that when people need to be accurate, and have enough time to do so, they will more often rely on perceptual simulation but that comes at a cost. Intuitively, in order to activate and process all those connected concepts, processing cannot be completed as quickly and so people rely on the linguistic associations when time is less available. These findings are compounded by the results of Experiment 2, which indicate that when participants make decisions about words presented in isolation they compare those words to the group of words included in the experiment. In these instances, utilizing distributional semantics from language statistics is a more efficient route. The Symbol Interdependency Hypothesis (Louwerse, 2007) argues that these symbols have been encoded with the grounded referents, so there is no need to activate all simulations when one symbol can easily and directly lead to another symbol. It stands to reason that over the course of a person's life where they make these symbolic connections over and over again, they can allow those shortcuts to make the connections for them. These findings support the view that language processing is both linguistic and embodied.

In a third experiment, I aimed to determine if both linguistic and perceptual representations are activated when generating relationships between word pairs. A RT experiment was designed where participants determined whether linguistically and/or perceptually similar or dissimilar word pairs were semantically related. As hypothesized, participants responded to linguistically related pairs with lower RTs than to pairs lacking such a relationship as a linguistic representation was sufficient to garner meaning. In addition, for word pairs that only shared a perceptual relation, a linguistic representation was not sufficient to determine the relationship between the word pair, and a perceptual representation was required. Importantly, unrelated words were processed in a similar fashion to words with a perceptual relationship, suggesting that participants generate a full perceptual representation when trying to relate dissimilar words.

As previously established in Experiment 1, linguistic features dominate early processing and perceptual features become salient later. It is unexpected then that participants responded to semantically related pairs with lower RTs than to pairs lacking such a relationship. As supported in prior work, such a finding implies that a quick linguistic representation was enough gather meaning for words that are frequently found in similar linguistic contexts. Furthermore, perceptually and semantically related word pairs were just as easy to process as pairs that were only semantically similar.

If the word pairs only shared a perceptual relation, then a linguistic representation was not sufficient in order to determine the relationship due to the lack of detailed perceptual information. Instead, in theory, a complete perceptual representation needed to be generated for both words to determine their perceptual similarities. Words pairs that were as unrelated as *strawberry* and *pony* were processed in a similar fashion to words a perceptual relationship. This finding helps to disentangle whether linguistic and perceptual representation are independent by demonstrating that with highly related word pairs, linguistic representations are good enough for generating meaning and with highly unrelated word pairs both linguistic and perceptual representations are necessary for comprehension.

Finally, linguistic factors predicted performance for semantically related pairs, whereas both linguistic factors and perceptual were necessary predict performance for perceptually related pairs. As expected for unrelated pairs, perceptual factors best predicted RT performance. These findings imply that a perceptual representation is necessary when trying to relate dissimilar words, but linguistic representations are more or less useless. The opposite is true for semantically related pairs, with linguistic factors predicting RTs and perceptual information remaining less than beneficial. Despite the relationship between the two factors, in this final chapter, the frequency pattern did not match the perceptual relation, suggesting that perceptual and linguistic systems are indeed acting independently.

Chapter Four

Language statistics and individual differences in processing primary metaphors

Abstract

Research in cognitive linguistics has emphasized the role of embodiment in metaphor comprehension, with experimental research showing activation of perceptual simulations when processing metaphors. Recent research in conceptual processing has demonstrated that findings attributed to embodied cognition can be explained through language statistics. The current chapter investigates whether language statistics explain processing of primary metaphors and whether this effect is modified by the gender of the participant. Participants saw word pairs with valence (Experiment 1: *good–bad*), authority (Experiment 2: *doctor–patient*), temperature (Experiment 3: *hot–cold*), or gender (Experiment 4: *male–female*) connotations. The pairs were presented in either a vertical configuration (*X* above *Y* or *Y* above *X*) matching the primary metaphors (e.g., HAPPY IS UP, CONTROL IS UP) or a horizontal configuration (*X* left of *Y* or *Y* left of *X*) not matching the primary metaphors. Even though previous research has argued that primary metaphor processing can best be explained by an embodied cognition account, results demonstrate that statistical linguistic frequencies also explain the response times of the stimulus pairs both in vertical and horizontal configurations, because language has encoded embodied relations. In addition, the effect of the statistical linguistic frequencies was modified by participant gender, with female participants being more sensitive to statistical linguistic context than male participants.

This chapter is based on:

Hutchinson, S., & Louwerse, M. M. (2013). Language statistics and individual differences in processing primary metaphors. *Cognitive Linguistics, 24,* 667 – 687.

Hutchinson, S., & Louwerse, M. M. (2012). The upbeat of language: Linguistic context and embodiment predict processing of valence words. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1709-1714). Austin, TX: Cognitive Science Society.

Introduction

Having argued in the previous chapter that the spatial location of stimuli affects how words are represented I provided an overview of some of the ways that linguistic and perceptual representations might be used differently in different situations. In the following pages I will discuss how the degree to which linguistic and perceptual information contribute to mental representations may even vary based, not only on the relative spatial location of the stimuli, but also upon the orientation of the stimuli. I will further discuss how individual differences (in this case gender) can impact how much participants rely on each type of representation when processing metaphoric language. In order to frame how these factors impact mental representations, I will present four experiments whereby I focus on how much the linguistic and perceptual factors explain male and female participant response times to valence words, authority words, temperature words, and gender words.

The current chapter has three goals, (1) to determine if, in addition to embodied mechanisms, language statistics can also account for metaphor processing, (2) to explore how individual differences impact conceptual processing and (3) to establish how the orientation of word pairs might influence processing.

First, researchers have made the claim that processing metaphors is so fundamentally embodied that no other factors are expected to be relevant for comprehension. Studies in cognitive linguistics demonstrate that conceptual metaphors emphasize relations between concepts and physical properties. These metaphors are thought to be understood through their relationships to physical space (Gibbs 1994, 2006; Kövecses 2005; Lakoff and Johnson 1980; 1999). For instance, expressions like *I am feeling up* or *I am flying high* mark the primary metaphor HAPPY IS UP, with positive terms being associated with 'up' properties and negative terms with 'down' properties. Similarly, the primary metaphor CONTROL IS UP associates concepts related to power and authority to high vertical positions. For example, a *supervisor* might find that in his *lofty* position, he is on *top* of a situation with everything *under* control. Accordingly, being subordinate is associated with low vertical positions; a new employee is on the *bottom* of the totem pole and must look up to his superiors. The source of these spatial properties is embodied: "These spatial orientations arise from the fact that we have bodies of the sort we have and that they function as they do in our physical environment" (Lakoff and Johnson 1980: 14). That is, "human conceptual processing is deeply grounded in embodied metaphor" (Gibbs 2006: 122).

Although it may seem obvious that frequency-based effects might be expected based on the previous chapters, many researchers in cognitive linguistics support an embodied view, suggesting that a variety of metaphors are understood through perceptual simulations of the body within physical space (Gibbs 1994, 2006; Lakoff and Johnson 1980; 1999). For instance, Wilson and Gibbs (2007) found that when participants performed an action congruent with a metaphorical statement, they were faster to comprehend the statements than when statements were incongruent with an action. After performing or imagining a grasping motion, a participant would be faster to process the metaphorical phrase *grasp a concept* than when performing or imagining other incongruent motions (Wilson and Gibbs 2007). Such findings demonstrate that metaphor comprehension can be explained by bodily action. Systematic linguistic analyses show that numerous additional metaphors are also grounded in bodily mechanisms (e.g., hunger, pressure, temperature, space, and emotion) (Gibbs, Lima, & Francozo, 2004; Kövecses 1986; Nayak and Gibbs 1990). For instance, abstract metaphors such as those linked to time seem to be processed through embodied mechanisms related to space. For example, the past is likely to be thought of as being behind us, as evidenced by phrases like *back in the day* or *that day is behind us now,* whereas the future is thought of as being ahead of us with phases like *are you looking forward to it* ? (Gibbs, 2006; Lakoff & Johnson, 1999.) In a number of studies, Boroditsky and colleagues demonstrated that time is indeed grounded in space with participants being likely to organize time spatially, such that spatial representations and manipulations influence how participants process and conceptualize time (Boroditsky, 2000, 2001; Boroditsky & Ramscar 2002). For example, participants moving forward in space are likely to think of time in terms of the TIME PASSING IS A MOVING OBSERVER metaphor (as opposed to the TIME PASSING IS A MOVING OBJECT metaphor) because they themselves are physically moving for ward in space. In addition, participants reading sentences describing motion in a particular direction are faster to process such sentences after witnessing motion in that same direction than after witnessing motion in a different direction (Dils & Boroditsky 2010).

The association between valence or authority concepts and specific physical locations can also be readily explained by embodied cognition. For example, evidence for the primary metaphor HAPPY IS UP comes from experiments that show that when an image or word representing a positive concept is presented on the top of a computer screen, comprehension is facilitated (Crawford et al., 2006; Meier & Robinson, 2005). For the primary metaphor CONTROL IS UP similar findings have been obtained. Studies have shown that when an authority word (e.g., *master*) was presented on the upper part of the screen, response times (RTs) were faster and recall was better than when a non-authority word (e.g., *servant*) was presented in that same position (Meier & Robinson, 2004; Meier, Bahrami, & Vigliocco, 2007; Schnall & Clore 2004; Schubert, 2005). In other words, comprehending primary metaphors such as HAPPY IS UP and CONTROL IS UP seems to be facilitated through the perceptual simulation of upward/downward directions, supporting the idea that primary metaphors are understood through their relationship to physical space.

In fact, results similar to the aforementioned primary metaphor experiments have been obtained using concepts with literal upward and downward locations, such as flying and swimming animals presented in the upper or the lower part of the screen (Pecher et al., 2010; Šetić & Domijan, 2007). Similarly, when word pairs with a perceptual order such as *attic* and *basement* are presented to participants, one above the other, iconic presentations are processed faster than reverse-iconic presentations (Zwaan & Yaxley, 2003). Such iconicity findings have also been obtained with other object words (Estes et al., 2008) and motion verbs (Meteyard et al., 2007). Over the last decade findings like these have accumulated, lending support to the claim that linguistic symbols are grounded in modality specific perceptual and motor systems (de Vega et al., 2008; Pecher & Zwaan, 2005; Semin & Smith, 2008). However, although conceptual metaphors are grounded in perception and action, perceptual simulations of the external world are not requisite for under standing and thinking about these metaphors (Boroditsky & Ramscar, 2002).

Contrary to this claim that metaphor comprehension is embodied only, Boroditsky and Ramscar (2002) and Gibbs (2006) have stated that embodiment effects do not account for the entirety of metaphor comprehension, but other factors are likely relevant too. One such candidate is language itself. A number of theories have cautioned against a unilateral embodied account of cognition. For instance, language has encoded up and down relations by typically placing the *up* concept before the *down* concept, as is the case in many binomials (e.g., it is more common to say *up and down*, *head to toe*, *top and bottom* than *down and up*, *toe and head*, and *bottom and top*) (Cooper & Ross 1975). That is, perceptual cues are encoded linguistically, such that language users can rely on the linguistic system as a shortcut to the perceptual system. Consequently, comprehension can be explained both by a statistical linguistic approach and a perceptual simulation approach. For instance, Louwerse (2008) demonstrated that findings of iconicity attributed to perceptual simulations, such as the facilitative iconicity effect when the word *attic* is placed above *basement*, could also be explained by statistical linguistic frequencies (the word order *attic–basement* is more frequent than the order *basement–attic*). These findings demonstrate that individuals rely on perceptual and linguistic information to varying degrees, based on a number of factors.

Following the Symbol Interdependency Hypothesis, if language encodes perceptual information, statistical linguistic frequencies should also be able to explain conceptual processing (Louwerse, 2008; Louwerse & Jeuniaux, 2010). In the past chapters, findings could be attributed to perceptual simulation, and also to statistical linguistic frequencies. However, all of these studies use stimuli found in the physical world, such as animals, numbers, objects (Louwerse, 2008; Louwerse & Jeuniaux, 2010), or modalities (Louwerse & Connell, 2011). Whether the same applies to non-literal relations, such as metaphors, is an open question. The experiments explore the possibility that metaphorical and literal thoughts are processed similarly enough, such that statistical linguistic frequencies are not only important during literal language comprehension but also when processing primary metaphors. Findings that have been predominantly attributed to perceptual simulations, such as the faster processing when authority words are positioned above non-authority words compared to the reverse, are predicted to also be attributed to statistical linguistic frequencies. Therefore, although embodied simulations account for a significant portion of metaphor comprehension, perhaps other factors also play an important role (Boroditsky & Ramscar, 2002; Gibbs, 2006; Louwerse, 2007, 2008, 2011).

Second, concerning individual differences, the question can be raised to what extent different participants are more or less accustomed to relying upon statistical linguistic frequencies when processing conceptual metaphors. Studies have shown that the relative prominence of statistical linguistic frequencies or perceptual simulation is modulated by the stage of conceptual processing (Louwerse & Connell, 2011), the cognitive task (Louwerse & Jeuniaux, 2010), and the stimulus (Louwerse & Jeuniaux, 2010). Yet it is unclear whether individual differences also modulate the relative importance of linguistic context. For instance, those participants with enhanced language skills may show an inclination to process information in a statistical linguistic fashion. This possibility can be explored through gender differences. Males tend to show greater general spatial ability and females show greater general and spoken language ability (Bourke & Adams, 2011; Kimura, 2000; Kramer, Delis, Kaplan & O'Donnell, 1997; Linn & Peterson, 1985). A large body of research finds gender differences between language and between spatial ability. For example, males often outperform females on spatial tasks (Benbow & Stanley, 1983;

Casey, Nuttall, & Pezaris, 2001). Similarly, females are likely to have superior language skills for a variety of tasks such as verbal fluency tasks, semantic categorization tasks, and verbal memory tasks (Andreano & Cahill, 2009; Burman, Bitan, & Booth, 2008; Bornstein, Haynes, Painter, & Genevro, 2000), as well as language advantages throughout development (Wei, Lu, Zhao, Chen, Dong, & Zhou, 2012). In addition, males are more likely to develop language disorders (Liederman, Kantrowitz, & Flannery, 2005; Rutter, Caspi, Fergusson, Horwood, Goodman, Maughan, Moffitt, Meltzer, & Carroll, 2004). Based on the aforementioned tendencies in gender, because females may have a greater affinity to encode information linguistically, this affinity is hypothesized to generalize to a semantic judgment task such that linguistic factors would be hypothesized to better predict female RTs than male RTs.

Finally, if language statistics can explain conceptual processing of word pairs in their vertical configuration it would provide evidence that primary metaphors can be explained by language statistics and embodiment together. If language statistics also explains processing of word pairs in their horizontal configuration — a configuration that provides little support for an embodiment account — it will provide further evidence for the importance of language statistics during conceptual processing (Louwerse, 2008).

In four experiments these three questions were answered (i.e., that of the role of statistical linguistic context, the effect of gender, and the impact of word pair orientation). In each experiment male and female participants responded to word pairs with either valence (positive and negative), authority (superior and inferior), temperature (hot and cold), or gender (male and female) connotations presented in either a vertical or a horizontal configuration.

Experiment 1: Valence

In Experiment 1, participants were presented with word pairs that were opposites on a valence dimension (e.g., *happy–sad*) either in a vertical configuration (*happy* above/below *sad*) or horizontal configuration (*happy* left/right of *sad*). Of interest was the extent to which statistical linguistic frequencies explained RTs to the iconic and reverse iconic presentation of primary metaphor word pairs. In addition, the interaction between participant gender and language statistics was explored.

Participants

Seventy-four native English-speaking undergraduates at the University of Memphis (53 females) participated for extra credit in a Psychology course. Thirty-four participants (24 females) were randomly assigned to the vertical presentation condition and forty participants (29 females) were randomly assigned to the horizontal presentation condition.

Materials

The experiment consisted of 50 pairs of words that were opposites on a valence dimension (e.g., *happy–sad*). To avoid a response bias towards the iconicity of the word pairs, 100 filler items consisted of word pairs without a positive-negative relation (e.g., *blade–walnut*), with half of the pairs having a high semantic association and half having a low semantic association as determined by latent semantic analysis (LSA), a computational linguistic technique that measures the similarity in meaning between word pairs, but ignores an order relation (Landauer et al. 2007) (high semantic association: *coos* = .44; low semantic relation: *cos* = .18).

Procedure

Participants were asked to judge the semantic relatedness of word pairs presented on an 800 × 600 computer screen running EPrime software (Psychology Software Tools Inc., Pittsburgh, PA). Words were presented one above another in the vertical condition, and next to one another in the horizontal condition. Upon the presentation of a word pair, participants indicated whether the pair was related in meaning by pressing designated yes or no keys. All word pairs were randomly ordered for each participant to negate any order effects. To ensure participants understood the task, participants completed five practice trials before beginning the experimental task.

Results

RTs above 2.5 *SD* from the mean per condition, per participant, were removed from the analysis, affecting 4.17% of the RT data.

Statistical linguistic frequencies

As in the previous chapters and previous studies (Louwerse, 2008; Louwerse & Connell, 2011; Louwerse & Jeuniaux, 2010) statistical linguistic frequencies were operationalized as the log frequency of a–b (e.g., *happy–sad*) or b–a (e.g., *sad–happy*) order of word pairs. The order frequency of all word pairs within 3–5 word grams was obtained using the large Web 1T 5gram corpus (Brants and Franz 2006).

Statistical linguistic frequencies indeed confirmed the HAPPY IS UP metaphor, with positive-negative ordered word pairs having a higher frequency than negative-positive word pairs, $t(44)$, 10.81, $p < .001$, $M = 12.15$, $SD = 2.77$ vs. $M = 10.47$, $SD = 2.66$.

For both vertical and horizontal conditions, a mixed effect regression analysis was conducted on RTs with the linguistic frequency as a fixed factor and participants and items as random factors . The model was fitted using the restricted maximum likelihood estimation (REML) for the continuous variable (RT). F-test denominator degrees of freedom were estimated using the Kenward Roger's degrees of freedom adjustment to reduce the chances of Type I error (Littell et al., 2002).

For the vertical configuration of word pairs, statistical linguistic frequencies explained RTs, $F(1, 84.55) = 58.80$, $p < .001$, $\eta^2 = .41$ with higher frequencies yielding lower RTs. For the horizontal condition, the statistical linguistic frequencies again explained RTs, $F(1, 86.21) = 82.44$, $p < .001$, $\eta^2 = .$

49, again with higher frequencies yielding lower RTs. These findings show that the HAPPY IS UP metaphor is encoded in language and that statistical linguistic frequencies explain the processing of this primary metaphor, both in vertical and horizontal configurations of the word pairs.

Participant gender effects

Having demonstrated that statistical linguistic context explained RTs, the next question to address was whether the effect was modulated by participant gender. To account for the differences in number of males (n = 21) versus

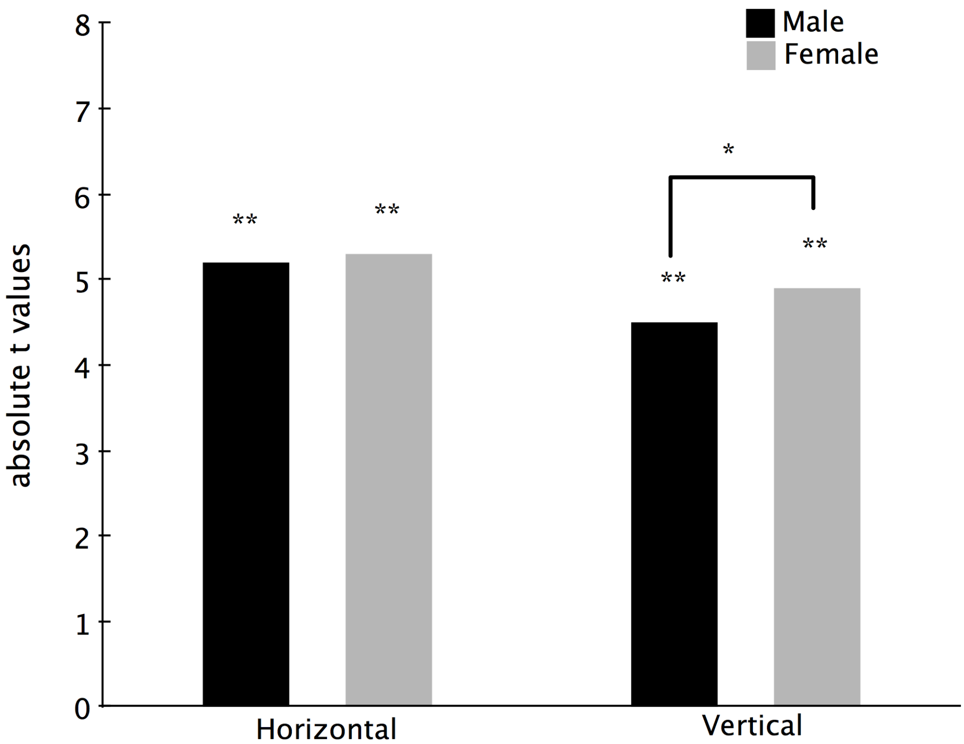


Figure 1. Male and female t-values from mixed effects analyses on RT from MCMC generated values in vertical and horizontal conditions of valence word pairs. Larger t-values indicate greater reliance on the linguistic factor. (* denotes significant effects at $p < .05$, ** denotes significant RT effects at $p < .001$)

females ($n = 53$), Markov chain Monte Carlo (MCMC) sampling generated 100 datasets with equal numbers of males and females. From this, 100 mixed effects analyses were conducted on RT for each gender with the linguistic variable as a fixed factor. Then these t-values were compared by gender. For the vertical condition, a t-test revealed that the t-values for each gender significantly differed, $t(198) = -4.80$, $p < .001$, with men having lower t-values than women (as linguistic frequency was more likely to predict female RTs than male RTs). However, for the horizontal condition, the t-values again were not significantly different between genders, $t(198) = -0.84$, $p = .40$ (Figure 1).

These results show the RT effects in both horizontal and vertical configuration of word pairs can be explained by statistical linguistic frequencies that encode the primary metaphor HAPPY IS UP. The effects are modulated by individual differences, with male participants relying less on the linguistic frequencies than female participants in the vertical condition, but not in the horizontal condition, where the vertical nature of such metaphors is not emphasized perceptually. Whether these findings can be extended to other metaphors is investigated in the next three experiments.

Experiment 2: Authority

Experiment 2 was identical to Experiment 1 except that it instead investigated whether statistical linguistic context determined processing of word pairs related to the Lakoff and Johnson (1999) CONTROL IS UP metaphor.

Participants

Seventy-nine different native English-speaking undergraduate at the University of Memphis (55 females) participated for extra credit in a Psychology course. Forty-one participants (28 females) were randomly assigned to the vertical condition and 38 participants (27 females) were randomly assigned to the horizontal condition.

Materials

The experiment consisted of 30 pairs of words that were opposites on an authoritative dimension (e.g., *parent–child*) and 60 filler items without an authority relation, with half of the pairs having a high semantic association and half having a low semantic association as determined by LSA (high semantic association: *cos* = .45; low semantic relation: *cos* = .20).

Procedure

The procedure was identical to Experiment 1.

Results

Subjects whose RTs fell more than 2.5 *SD* from the mean per condition, per participant were removed from the analysis, affecting 8.64% of the data.

Statistical linguistic frequencies

Statistical linguistic frequencies were operationalized in the same way as for Experiment 1. Again, evidence was found that the primary metaphor was encoded in language, with powerful–powerless word pairs being more frequent than powerless–powerful word pairs, $t(21) = 6.42$, $p < .001$, $M = 11.36$, $SD = 2.71$ vs. $M = 9.70$, $SD = 3.01$.

For both the vertical and horizontal configurations, mixed effect regressions were again conducted on RTs with linguistic frequency as the fixed factor and participants and items as random factors.

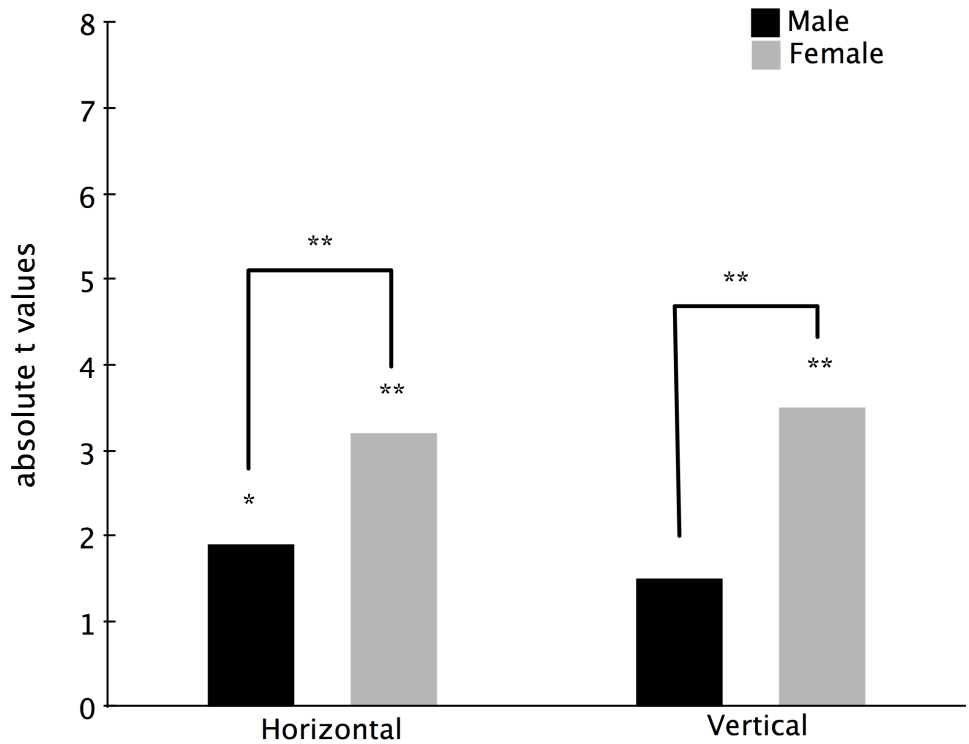


Figure 2. Male and female t-values from mixed effects analyses on RT from MCMC generated values in vertical and horizontal conditions of authority word pairs. Larger t-values indicate greater reliance on the linguistic factor. (* denotes significant RT effects at $p < .05$, ** denotes significant RT effects at $p < .001$)

As in Experiment 1, for the vertical condition linguistic frequencies explained the RTs, $F(1, 45.77) = 32.71$, $p < .001$, $\eta^2 = .42$, with higher frequencies yielding lower RTs. For the horizontal condition, the linguistic factor was also related to the RT, $F(1, 45.07) = 16.25$, $p < .001$, $\eta^2 = .26$, with higher frequencies yielding lower RTs. These findings again show that primary metaphors are encoded in language, and that both the vertical and horizontal configuration of conceptual metaphors can be explained by the linguistic system, confirming previous findings that statistical linguistic frequencies explain conceptual processing.

Participant gender effects

MCMC sampling generated 100 datasets with equal numbers of males and females. For the vertical configuration, a t-test revealed that the t-values for each gender significantly differed, $t(198) = 23.54$, $p < .001$, with women having higher t-values than men (as linguistic frequency was more likely to predict female RTs than male RTs). The same was found for horizontal condition, $t(198) = 12.41$, $p < .001$ (Figure 2). The findings of Experiment 2 with authority words extended those in Experiment 1 with valence words. The linguistic factor again explained RTs both in the vertical and horizontal conditions, with effect sizes similar to those found in Experiment 1. Furthermore, as expected, female participants relied more on the statistical linguistic factor than male participants in both vertical and horizontal orientations.

Experiment 3: Temperature

Lakoff and Johnson (1980) claim that heat is affiliated with an upward location, whereas cold is affiliated with a downward location by giving the example *If you're too hot, turn the heat down*. Similarly, IJzerman and Semin (2009, 2010) give the example that *holding warm feelings toward someone* or *giving someone the cold shoulder* can be associated with positive (up) and negative (down) social relations, respectively. Furthermore, examples like *she was boiling with anger* suggest that an increase in body temperature is physiologically linked to anger and also the associated physical upwards motion (Pacini & Barnard, 2011). In Experiment 3 the finding for valence and authority was extended to temperature. That is, it was investigated whether hot and cold concepts can also be explained by statistical linguistic context, and whether these effects are modulated by participant gender.

Participants

Seventy-two different undergraduate native English speakers at the University of Memphis (53 females) participated for extra credit in a Psychology course. Forty-one participants (29 females) were randomly assigned to the vertical condition and 31 participants (24 females) were randomly assigned to the horizontal condition.

Materials

The experiment consisted of 22 pairs of words that were opposites on an temperature dimension (e.g., *hot–cold*) with 44 filler items without a temperature relation, with half of the pairs having a high semantic association and half having a low semantic association (high semantic association: *cos* = .49; low semantic relation: *cos* = .20).

Procedure

The procedure was identical to that of Experiments 1 and 2.

Results

Subjects whose RTs fell more than 2.5 *SD* from the mean per condition, per participant were removed from the analysis, affecting 4.30% of the data.

Statistical linguistic frequencies

The operationalization of statistical linguistic frequencies was the same as for Experiments 1 and 2. Following the findings in the previous experiments that primary metaphors are encoded in language, hot–cold word pairs were significantly more frequent than cold–hot word pairs, $t(17) = 6.12$, $p < .001$, $M = 11.55$, $SD = 3.40$ vs. $M = 10.67$, $SD = 3.37$.

For both conditions, the same mixed effect analysis was conducted on RTs. As in Experiment 1 and 2, for the vertical condition the linguistic factor explained the RTs, $F(1, 33.16) = 31.31$, $p < .001$, $\eta^2 = .49$, with higher frequencies yielding lower RTs. For the horizontal condition, the linguistic factor also predicted RTs, $F(1, 33.51) = 30.63$, $p < .001$, $\eta^2 = .48$, with higher frequencies yielding lower RTs. These findings are similar to those obtained in the vertical condition, demonstrating that participants relied on statistical frequencies when processing word pairs related to temperature. Specifically, those word pairs matching the primary metaphor were processed faster than their counterparts, again both in vertical and horizontal configurations.

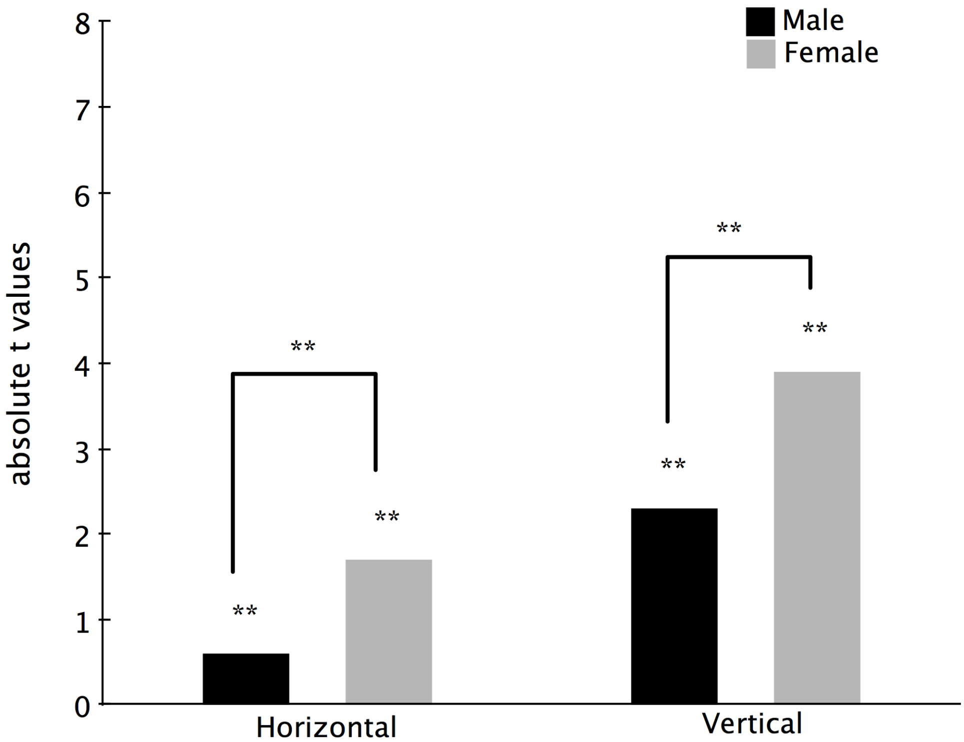


Figure 3. Male and female t-values from mixed effects analyses on RT from MCMC generated values in vertical and horizontal conditions of temperature word pairs. Larger t-values indicate greater reliance on the linguistic factor. (** denotes significant RT effects at $p < .001$)

Participant gender effects

As in Experiment 1 and 2, MCMC sampling generated 100 datasets to account for differences in number of males and females. A t-test revealed that the t-values for each gender significantly differed in the vertical condition, $t(198) = 11.14$, $p < .001$, with women having higher t-values than men (as linguistic frequency was more likely to predict female RTs than male RTs). For the horizontal condition, similar results were obtained, $t(198) = 7.23$, $p < .001$ (Figure 3). The findings of Experiment 3 with temperature words replicated those in Experiment 1 and 2. The linguistic factor explained RTs for both configuration conditions, with effect sizes similar to those found in Experiments 1 and 2. Again, females relied more on the statistical linguistic factor than males in both conditions.

Experiment 4: Gender

Meier and Dionne (2009) extended the CONTROL IS UP metaphor to gender, whereby males are (stereotypically) considered to be powerful and females are (stereotypically) considered to be powerless. The prediction would then be that male concepts placed in upward physical location are processed faster than female concepts in that location. Meier and Dionne (2009) indeed demonstrated that pictures of female faces were thought to be more attractive when presented at the bottom of a screen (a location–concept match), whereas male faces were thought to be more attractive when presented at the top of a screen. Experiment 4 investigated word pairs such as *male–female* and *female–male* in vertical and horizontal configurations, similar to Experiments 1–3 investigating valence, authority, and temperature.

Participants

Seventy-seven different undergraduate native English speakers at the University of Memphis (54 females) participated for extra credit in a Psychology course. Thirty-nine participants (27 females) were randomly assigned to the vertical condition and 38 participants (27 females) were randomly assigned to the horizontal condition.

Materials

The experiment consisted of 32 pairs of words that were opposites on a gender dimension (e.g., *male–female*). Sixty-four filler pairs lacked a gender relation, with half of the pairs having a high semantic association and half having a low semantic, (high semantic association: *cos* = .48; low semantic relation: *cos* = .22).

Procedure

Procedure was identical to that of Experiments 1–3.

Results

RTs that fell more than 2.5 *SD* from the mean per condition, per participant were re moved from the analysis, affecting 2.45% of the data.

Statistical linguistic frequencies

Linguistic frequencies were operationalized in the same way as Experiments 1–3. Male–female pairs had higher frequencies than female–male pairs, $t(31) = 7.42$, $p < .001$, $M = 10.72$, $SD = 4.13$ vs. $M = 9.46$, $SD = 3.86$.

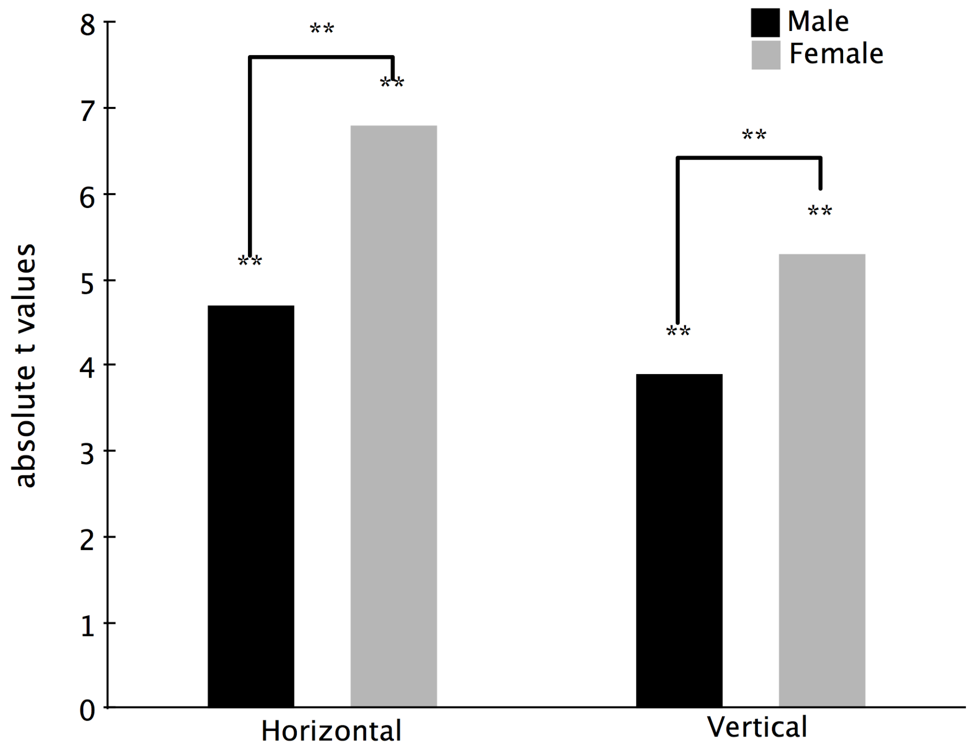


Figure 4. Male and female t-values from mixed effects analyses on RT from MCMC generated values in vertical and horizontal conditions of gender word pairs. Larger t-values indicate greater reliance on the linguistic factor. (** denotes significant RT effects at $p < .001$)

For both conditions, the same mixed effect analysis was conducted on RTs as before. For the vertical condition statistical linguistic frequencies explained RTs, $F(1, 60.64) = 35.43$, $p < .001$, $\eta^2 = .37$, with higher frequencies yielding lower RTs. Linguistic frequencies also predicted RTs in the horizontal configuration, $F(1, 59.37) = 46.55$, $p < .001$, $\eta^2 = .44$, with higher frequencies yielding lower RTs. These findings are similar to those obtained in the other experiments and support the conclusion that processing of gender concepts in vertical and horizontal configurations can be explained by statistical linguistic frequencies.

Participant gender effects

As before, it was tested whether linguistic factors better predicted RTs for females. For the vertical condition, a t-test revealed that the t-values for each gender significantly differed in the vertical condition, $t(198) = 26.21$, $p < .001$, with women having higher t-values than men. For the horizontal condition, similar results were obtained, $t(198) = 30.51$, $p < .001$ (Figure 4). The findings of Experiment 4 with gender words replicate those in the previous Experiments. The linguistic factor explained RTs for both configuration conditions, with effect sizes similar to those found in Experiments 1 and 2. Furthermore, females relied more on the statistical linguistic factor than males in both conditions.

Conclusion

In the previous chapter, four experiments made it clear that primary metaphor processing can best be explained by both an embodied cognition account and a linguistic account. Importantly, whether the stimuli were presented in vertical or horizontal configurations modulates how much participants rely on linguistic representations, presumably because language has encoded embodied relations. Also important, the effect of the statistical linguistic frequencies was modified by participant gender, with female participants being more sensitive to statistical linguistic context than male participants, thereby demonstrating that individual differences as well as stimuli orientation can impact whether participants rely more on embodied or linguistic representations.

These four experiments investigated whether valence (e.g., *good–bad*), authority (e.g., *doctor — patient*), temperature (e.g., *hot–cold*), and gender (e.g., *male–female*) word pairs with matched iconicity were processed faster than mismatching word pairs. This investigated to what extent statistical linguistic frequencies and perceptual factors explained RTs of word pairs related to valence (Experiment 1), authority (Experiment 2), temperature (Experiment 3), or gender (Experiment 4). The findings were not only considered in terms of significance, but also in terms of effect size. For example, Meier and Robinson (2004) obtained significant results, with positive words being processed faster on the top of the screen, and with negative words being processed faster on the bottom of the screen. Although this effect clearly supports an account of embodied cognition, the authors do not address the effect sizes associated with their results (R^2 = .16 for Experiment 1, and R^2= .23 for Experiment 2). The effect sizes for the statistical linguistic frequencies reported in this chapter were consistently at least two times larger than effect sizes obtained from previous embodied cognition research such as Meier and Robinson (2004). Finally, different from most embodied cognition studies RTs were analyzed using a mixed effects model with both participant and items as random factors. Variance attributed to different participants and, importantly, variance attributed to different items was removed from the analysis to avoid the language-as-fixed fallacy effect, thereby yielding more reliable results (Clark, 1973; Brysbaert, 2007). This issue is discussed further in Chapter 6.

Do the results reported here suggest that statistical linguistic frequencies should be seen as an explanation for conceptual processing in lieu of perceptual simulation? Certainly not, there is no denying that embodiment offers some explanatory power during metaphor comprehension, however it is not the only important factor at play. There are several reasons embodiment should not be overlooked.

In fact, some metaphors like these (e.g., HAPPY IS UP) seem to be near universal, and this makes sense if metaphor comprehension is based on bodily perception and action. At the same time, many conceptual metaphors at the specific level vary cross culturally and even regionally, suggesting that embodiment cannot entirely account for metaphor comprehension on its own (Kövecses, 2005). In point of fact, cultures even vary with how they represent spatial information with relationship to their bodies, raising the question of how language comprehension can rely only on the body, when different cultures think of their bodies as being related to space in different ways. Kövecses (2005) argues that the body does not exist in isolation, instead it is molded by our environment, our history, and our interests thus introducing variance in how language is understood through the body. In addition to these factors, language itself influences metaphor comprehension, with statistical linguistic frequencies explaining RTs in the current experiments. But these findings beg the question of why it is the case that *happy–sad* (*hot–cold*, *teacher–student*, *male–female*) is more frequent than the reverse word pair orders? The answer is that language encodes perceptual relations. That is, when speakers formulate utterances, pre linguistic conceptual knowledge is translated into linguistic conceptualizations, so that as a function of language use, perceptual relations become encoded in language (Louwerse, 2008). As shown in the current chapter, antonyms of valence, temperature, authority, and gender demonstrate that an embodied cognition account is not detached from a linguistic account. On the contrary, a statistical linguistic account complements an embodied cognition account (Louwerse, 2011). The linguistic system has evolved to encode perceptual relations and these linguistic cues are employed by language users during their cognitive processes, as demonstrated in both vertical and horizontal configurations of word pairs.

In addition to investigating the role of statistical linguistic frequencies on conceptual processing, this chapter investigated to what extent male and female participants were affected by these two factors differently. Prior studies have investigated whether stimulus (word or picture), cognitive task (semantic judgment or iconicity task), or stage of processing (faster or slower RTs) affected the role of statistical linguistic frequencies and perceptual simulation differently. The current chapter investigated whether participant gender could be added to the series of modulators of statistical linguistic frequencies and conceptual processes. The findings in this chapter demonstrate that gender should indeed be considered a modulator. These experiments demonstrated that female participants typically relied more on statistical linguistic patterns. However, it should be noted that although it can be speculated that these findings might be due to differences in linguistic and spatial ability between genders such constructs were not directly measured. Such a step would be useful in future research to ensure such effects are indeed driven by linguistic and spatial ability and are not simply due to other factors linked to gender.

Finally, the orientation of word pairs in an experimental setting further suggests that embodied and linguistic representations are both relevant to varying extents. While language statistics can explain conceptual processing of word pairs in their horizontal configuration therefore providing further evidence for the importance of language statistics during conceptual processing (Louwerse, 2008), it is hard to understand how an embodied relationship could exist between word pairs presented horizontally. Similarly when presented in a vertical condition where the vertical nature of such metaphors is emphasized perceptually, statistical linguistic frequencies account for less variance in RTs. This is a logical result, assuming that prior work is correct in the conclusion that embodied representation account for facilitated responses for pairs presented in their expected iconic ordering.

The findings of these experiments show that even though perceptual simulations contribute to RT differences, it is important to consider the extent to which other factors, such as language statistics, explain processing. The mind is embodied but is also linguistic, while participant gender, and stimuli orientation determines how much.

Chapter Five

Effect size matters: The role of language statistics and perceptual simulation in conceptual processing

Abstract

The cognitive science literature increasingly demonstrates that perceptual representations are activated during conceptual processing. Such findings suggest that the debate on whether conceptual processing is predominantly symbolic or perceptual has been resolved. However, studies too frequently provide evidence for perceptual simulations without addressing whether other factors explain dependent variables as well, and if so, to what extent. The current paper examines effect sizes computed from 126 experiments in 51 published embodied cognition studies to clarify the conditions under which perceptual simulations are most important. Results showed that effects of language statistics tend to be as large or larger than those of perceptual stimulation. Moreover, factors that can be associated with immediate processing (button press, word processing) tend to reduce the effect size of perceptual simulation. These findings are considered in respect to the Symbol Interdependency Hypothesis, which argues that language encodes perceptual information, with language statistics explaining quick, good-enough representations and perceptual simulation explaining more effortful, detailed representations.

This chapter is based on:

Louwerse, M. M., Hutchinson, S., Tillman, R., & Recchia, G. (2014). Effect size matters: the role of language statistics and perceptual simulation in conceptual processing. *Language,Cognition, and Neuroscience.*

Introduction

In the previous chapters I concluded that linguistic and perceptual representations are independent but highly related processes that are both relied upon in different ways, at different times, to different extents during language processing. Although these findings support integrated theories of cognition the debate on whether conceptual processing is predominantly symbolic or perceptual is still an active area of research. In the following chapter I will provide an overview of studies that provide evidence for perceptual simulations without addressing other factors. This chapter examines effect sizes computed from 126 embodied cognition experiments to establish the conditions under which perceptual simulations versus linguistic representations are most important.

The conclusion in most embodied cognition studies that perceptual simulation is activated in conceptual processing assumes a one-size-fits-all approach. It seems plausible, considering the evidence reported in the previous chapters, that the effect of perceptual simulation is modulated by the experimental variables being used. The question of whether conceptual processing activates perceptual simulations can then be replaced by the question of under what conditions perceptual simulations are most active. That question seems to be more productive, as it considers the extent various factors, including perceptual simulation and language statistics factors, play a role, rather than simply asking if these factors play a role. In the previous chapters and in previous work, I have attempted to answer that question, by considering factors like the time course of processing, the spatial presentation of stimuli, individual differences, and the orientation of stimuli.

Nature of the stimulus

Barsalou (1999) argued convincingly that it is unlikely that processing a picture will simply generate an amodal symbolic representation. Instead, perceptual processes lead to perceptual representations. Others would agree with this view. For instance, Kintsch (1998, p. 47) stated that perceptual symbols, imagery and actions are among the building blocks of cognition, and Landauer and Dumais (1997, p. 235) argued that perceptual world knowledge underlies an associative learning theory. The important question, however, is whether linguistic symbols must always be transduced into perceptual symbols. Barsalou (1999, p. 652) acknowledged the importance of structured representations, propositions, frequency effects, and pattern completion in conceptual processing. Similarly, Paivio's (1971, 1986) Dual Coding Theory postulated that both visual and verbal information are processed differently and along distinct channels. Others strongly argued that linguistic symbols must be transduced into perceptual simulations (Pecher & Zwaan, 2005), leaving little room for an amodal symbol system (Glenberg & Robertson, 1999).

In the previous chapters, I demonstrated that we interpret both linguistic and non-linguistic stimuli with both linguistic and non-linguistic processes, but the extent to which a particular type of process will dominate processing can be expected to depend on the nature of the stimuli (e.g., Louwerse & Jeuniaux, 2010). That is, pictures also activate linguistic representations and words also activate perceptual representations. Louwerse and Jeuniaux (2010) investigated this question by comparing the effect of both language statistics and perceptual ratings on the response times in a semantic judgement task of words and pictures. Concept pairs such as monitor and keyboard were presented in a vertical configuration. In one set of experiments, the concept stimuli were words; in the other the concept stimuli were pictures. Participants were asked to evaluate the relationship between the two concepts. The results showed that language statistics and perceptual simulations both explained the response times to both pictorial and linguistic stimuli, but language statistics explained word processing better than perceptual ratings did, but for pictures the effect of perceptual ratings dominated the effect of language statistics.

Many embodied cognition studies have shown that an independent variable associated with perceptual simulation explains response times. What is often unknown is what the effect size is of this association and whether another factor (e.g., language statistics) explains the response times equally well or – in the case of linguistic stimuli – better.

Individual differences

Do people activate perceptual simulations similarly? A few studies have provided insight for this question. Dijkstra et al. (2004) found that the effect of a picture mismatching a sentence was stronger for older than for younger adults. The authors explained this effect by older adults being involved in deeper processing, whereas younger adults focus more on the surface structure of the sentence, resulting in a weaker perceptual simulation effect. Holt and Beilock's (2006) finding that comprehenders with experience on a topic yield perceptual representations that are more differentiated than comprehenders without experience suggests a similar direction. Deeper understanding of the stimuli yields (deeper) perceptual simulations.

In Chapter 4, I showed that gender differences also impact how much participants rely on perceptual or linguistic representations in conceptual processing as men typically perform better on spatial and perceptual tasks (Benbow & Stanley, 1983; Casey, Nuttall, & Pezaris, 2001), whereas females typically outperform men on language tasks (Andreano & Cahill, 2009; Bornstein, Haynes, Painter, & Genevro, 2000; Burman, Bitan, & Booth, 2008). When making semantic judgements on word pairs, female participants relied more on statistical linguistic frequency patterns whereas males relied less so on linguistic representations. Such findings indicate that factors related to gender can also impact the extent to which individuals rely on statistical linguistic frequencies during language processing.

The effect of perceptual simulations and linguistic representations on conceptual processing seems to be modulated by individual differences, such as age or expertise. The effect size might even be affected by participant gender, making it a relevant factor to take into account in drawing conclusions on the nature of conceptual processing.

Cognitive task

The Dijkstra et al. (2004) and Holt and Beilock (2006) studies show that experience yields larger effect sizes in the relation between perceptual simulation and response times. It can thereby be argued that experience seems to yield deeper processing, and deeper processing yields perceptual simulations. Can deeper processing also be manipulated by the cognitive task participants need to perform? There is some evidence from the embodied cognition literature that this is the case. Borreggine and Kaschak (2006) asked participants to hold a keyboard on their lap in a 90° angle, so that the Q key was located near the participant and the P key away from the participant. They then listened to a sentence like *Joe kicked you the soccer ball* or *You kicked Joe the soccer ball*. Participants were asked whether the sentence was sensible. Near-body responses yielded faster responses for *Joe kicked you the soccer ball* with an opposite effect for *You kicked Joe the soccer ball* and reverse for the away-from-body responses. When the participant was able to prepare and execute the motor response, the effect was strongest. Knowing what motor action was required elicited an action–sentence compatibility effect, perhaps because of the deep conceptual processing of the stimuli.

Louwerse and Jeuniaux (2010) tested the effect of the cognitive task on the extent to which language statistics or perceptual ratings predicted RT values. Participants were asked to either make semantic judgements or iconic judgements regarding words that shared iconic and semantic relationships. That is, for semantic judgements participants were asked to determine whether the concept pairs had a similar meaning, whereas for the iconicity judgements participants were asked whether the concepts had the same relation in the perceptual world, as the relation presented on the screen (e.g., monitor above keyboard). Cognitive processing in the iconicity judgement task was deeper than in the semantic judgement task, because a prerequisite for the iconicity judgement was a semantic judgement. Effect sizes for perceptual simulation were larger for the deeper cognitive task (iconicity judgement) than the shallow cognitive task (semantic judgement) with the opposite result for effect sizes for language statistics.

These findings suggest that when participants form detailed, full-fledged representations of the concepts being presented to them, because they have the experience to do so, because they plan their response or because the cognitive tasks requires them to do so, the effect size of perceptual simulation seems to be larger.

Time course of cognitive processing

If language statistics dominate in shallow cognitive tasks and perceptual simulation dominates in deeper cognitive tasks, and if deeper cognitive tasks imply shallow cognitive tasks, the prediction is that in conceptual processing language statistics precede perceptual simulation. This is what the Language and Situated Simulation (LASS) theory predicts (Simmons, Hamann, Harenski, Hu, & Barsalou, 2008). In conceptual processing, both the linguistic system and the simulation system become active initially, but activation in the linguistic system peaks first. In an fMRI experiment, Simmons et al. (2008) found that activations early in processing overlapped with activations for word associations (Broca's area in the left inferior frontal gyrus), whereas activations late in processing overlapped with activations for situation generation (right posterior Superior Temporal Sulcus). "When linguistic forms and associated statistical information are sufficient for adequate performance, no retrieval of conceptual information is necessary. This does not mean that these strategies are insignificant, given their obvious heuristic value" (p. 107). The Symbol Interdependency Hypothesis (Louwerse, 2007; Louwerse & Jeuniaux, 2008) makes the same prediction. Because the linguistic system encodes perceptual information, language statistics allows for good-enough representations, whereas perceptual simulation allows for deeper conceptual representations. Louwerse and Connell (2011) tested this on modality words, in both a computational linguistic and an experimental setting. The computational linguistic analysis showed categorizing modality words based on linguistic frequency resulted in three categories: visual/haptic, auditory and olfactory/gustatory. Whereas a perceptual account clearly distinguishes between five modalities, the language statistics account only allowed for a less fine-grained classification. They predicted that if language statistics dominate early in conceptual processing, the effect size of language statistics should be highest for fast response times and lowest for slow response times, whereas the opposite should hold true for perceptual simulations. In a response time experiment, they found that the coarse- grained language statistics variable best explained fast response time and fine-grained perceptual simulation variable best explained slow response times, with both variables explaining medium response times.

Louwerse and Hutchinson (2012) extended this conclusion in an electroencephalography (EEG) experiment. In a task where subjects made semantic or iconic judgements about word pairs, neural activity over the time course of each trial was recorded. The objective was to compare activity, over the trial's duration, between regions of the brain either commonly associated with linguistic processing or with perceptual processing. They found linguistic cortical regions were relatively more active than perceptual cortical regions early in a trial. However, the reverse was true later in the trial, supporting the notion that language statistics dominate in early processing and perceptual representations become more important as a trial progresses.

Linguistic forms and associated statistical information allow for heuristic processing (Simmons et al., 2008), fast, good-enough representations (Louwerse & Connell, 2011; Louwerse & Hutchinson, 2012), whereas perceptual simulation allows for a deeper conceptual understanding. Another way to think about the relationship between shallow and deep processing is by analogy to the distinction between “System 1” and “System 2” (Stanovich & West, 2000) popularized by Kahneman (2003).

Kahneman does not support a modular view of mind and emphasizes the two systems are not actual systems but are merely used for illustrative purposes. System 1 refers to a set of cognitive processes that occur quickly, automatically, in parallel, and below the level of conscious awareness, while System 2 refers to more effortful, controlled, sometimes rule-governed processes that generally involve the deployment of limited, cognitively expensive resources such as attention and willpower. While some embodied processes certainly belong to System 1 as well – particularly processes involving associations between words and emotions, or words and basic percepts – it is unlikely that language users construct a complete mental simulation for every sentence they comprehend (Louwerse, 2011). Perceptual simulations that go beyond simple association, therefore, are more in the domain of System 2: too resource-intensive to be undertaken in circumstances under which associations among words will suffice for the completion of the task at hand (Louwerse & Connell, 2011).

In conclusion, factors such as stimulus and cognitive task modulate the effect of perceptual simulation on conceptual processing. What is unknown, however, is to what extent modulators such as stimulus and cognitive task affect the results of published embodied cognition studies.

Effect size matters

Perceptual Simulation

Knowing the estimated magnitude of results reported in published embodied cognition studies would allow for (1) measuring whether the average effect size of perceptual simulation variables is small or large and (2) determining factors which influence the effect size. The problem with an investigation like this is that the modulators identified earlier are often not specified in embodied cognition studies. The following overview of embodied cognition studies below should therefore be seen as an exploration.

Searches for the studies to be included in the analysis were conducted between 2010 and 2012. Search terms for PsycInfo and Google Scholar were embodiment, embodied cognition and perceptual simulation. Fifty-one studies were extracted on the basis of the following criteria: (1) studies had to address some component of embodiment (e.g., action compatibility, spatial orientation), (2) be published between 1999 and 2012 in a peer-reviewed journal with an impact factor >1 (M = 3.6, SD = 1.46), (3) include experiments; theoretical papers were not included; (4) include language stimuli; gesture studies or studies using only pictorial or video stimuli but not language were excluded; (5) include statistically analyzable and comparable results, such as ANOVAs or mixed effects models, and their details, such as *F* values, *df*'s, *M*, and *SD*.

From the 126 experiments in these 51 studies, information was extracted regarding the type of response (e.g., response time or rating), the method of response (e.g., button press or otherwise), stimulus type (e.g., pictures, words), and when relevant to the experimental design, information regarding the domain of stimuli (e.g., configuration of the stimuli) was also collected.

For each experiment, *df*, *t*, *N* and *F* values were collected from critical analyses demonstrating embodied effects. Because only 33 out of the 126 experiments (27%) reported effect sizes, the variance explained (partial η^2 or R^2 values) for each experiment was computed using the method in Fritz, Morris, and Richler (2012) and Louwerse and Jeuniaux (2010). That is, the strength of a model association is represented as a weighted ratio of the *F* statistic. R^2 and *F* used in ordinary regression analysis are closely related, since $F = (R^2/k) / ((1 - R^2) / (N- k - 1))$ where *k* is the number of model parameters and *N* is the number of cases, such that *F* has $(k, N - k - 1)$ *df*. See also Pedhazur (1997, p. 105). Thus, partial eta squared (η_p^2) can be calculated as $(df_{\text{effect}} \times F_{\text{effect}}) / ((df_{\text{effect}} \times F_{\text{effect}}) + df_{\text{error}})$. This formula represents the sum of squares of the effect of interest, divided by the sum of squares of the error plus the sum of squares of the effect. This means that for any experiment that reports *df* and *F* values, an effect size of partial eta squared can be calculated.

The average effect size of the 126 experiments was η^2 = .185 (*SD* = .177). According to Cohen (1988, p. 287), this can be viewed as a large (η^2 > .14) effect size, explaining approximately 19% of the variance of the dependent variable.

Next I aimed to determine which factors best explained the effect sizes. Four variables occurred frequently in the selected studies: the use of response time as the dependent variable, the use of vertical configuration of the stimuli, button press as the method of response and the use of single words rather than sentences as stimuli (Appendix 1). I dummy-coded (Cohen & Cohen, 1983) stimulus type (i.e., single words), type of response (i.e., RT), method of response (i.e., button press) and domain of stimuli (i.e., vertical configuration), and ran a mixed effects regression model using the four dummy-coded variables as independent variables and the effect sizes as dependent variables. Whether or not a response time method was used did not affect the η^2 effect size, $F(4,121)$ = 1.82, p = .18, η^2 = .057, (*M* = .194, *SD* = .134 vs. *M* = .182, *SD* = .185) and neither did the vertical configuration of the stimuli, $F(4, 121)$ = .536, p = .47, η^2 = .017,(*M* = .184, *SD* = .178 vs. *M* = .184, *SD* = .176). Whether or not single words (as opposed to sentences, paragraphs or pictures) were used as stimuli did influence η^2, with the use of single words reducing effect size, $F(4, 121)$ = 3.52, p = .04, η^2 = .104, (*M* = .162, *SD* = .144 vs. *M* = .227, *SD* = .221). Similarly, whether or not participants were asked to respond to stimuli with button presses influenced η^2 with the use of button presses marginally reduced the effect size,

$F(4, 121) = 4.29$, $p = .06$, $\eta^2 = .124$, ($M = .140$, $SD = .111$ vs. $M = .197$, $SD = .190$).

Computing the effect size η^2 is useful, because the cognitive science community is most familiar with this measure of effect size. However, the use of η^2 comes at a price. Fritz et al. (2012) caution that η^2 as an effect size value becomes less meaningful when comparing different studies with different error terms. In this chapter, I therefore also calculated the less common but in case of different error terms more accurate Hedge's *g* effect size. However, some experimental designs are not optimal for computing Hedge's *g* (e.g., three or more groups, unspecified number of subjects), so both measures of effect size were included.

Hedge's *g* is a measure of effect that does not rely on *df* but instead is calculated as $(t \times \sqrt{((1/n_1)+(1/n_2))} \times (1 - (3/(4 \times (n_1+n_2) - 9)))$ for between subject designs and as $((2 \times t)/\sqrt{n}) \times (1 - (3/(4 \times n - 9)))$ for within subject designs. The average effect size of Hedge's *g* was = .965 ($SD = .711$), considered a large effect (Lakens, 2013). The correlation between Hedge's *g* measure of effect and η^2 was strong at $r\ (126) = .857$.

However, performing the same analysis now using Hedge's *g*, only the use of single words explained the dependent variable of effect size, $F(4, 121) = 10.452$, $p = .002$, $\eta^2 = .257$, ($M = .826$, $SD = .438$ vs. $M = 1.224$, $SD = 1.002$). Type of response (e.g., RT) $F(4, 121) = .384$, $p = .536$, $\eta^2 = .013$, ($M = .898$, $SD = .387$ vs. $M = .976$, $SD = .762$), stimuli configuration $F(4, 121) = 2.31$, $p = .$

131, η^2 = .071, (M = .982, SD = .731 vs. M = .892, SD = .626) and method of response (e.g., button press) F(4,121) = 2.35, p = .128, η^2 = .072, (M = .795, SD = .310 vs. M = 1.011, SD = .781) did not explain the effect sizes, even though the patterns were the same as with η^2.

These results suggest that if the stimulus is simple, for instance a word, the effect sizes supporting an embodied cognition account are smaller than those responses that move beyond button presses or word level stimuli. When responses are quick, for instance by means of button presses, there are indications that effect sizes may be smaller, but this was not confirmed by a Hedge's *g* analysis.

There are a number of explanations for the finding that the use of single words reduces the effect size in embodied cognition findings. One explanation is that responses to simple stimuli such as words are noisier than responses that are more complex, such as sentences. That is, without being constrained by context single words exhibit greater variance than responses to sentences, which can likely be explained by single words being semantically noisier than context-specific sentences (see for an identical argument that responses to individual sentences are noisier than combinations of sentences Yang, Mo, and Louwerse, 2012).

Another explanation is that larger text units such as sentences or paragraphs require deeper processing and higher engagement and hence yields larger effect sizes. The smaller effect sizes for single word stimuli would then be explained by the cognitive task. Indeed, Louwerse and Jeuniaux (2010) have demonstrated that a deeper cognitive task, one that implies another task (e.g., an iconicity task that implies a semantic judgement task) yields a larger effect size for a perceptual simulation factor than for language statistics factor, whereas a relatively shallow cognitive task yields a larger effect for a language statistics factor than a perceptual simulation factor.

A third explanation focuses on the time course of processing: single words yield quicker processing and quicker processes are better explained by language statistics than by perceptual simulation. This explanation is supported by Louwerse and Connell (2011) who showed that slower response times are best explained not only by perceptual simulation, but also by Louwerse and Hutchinson (2012) who showed that perceptual cortical regions compared to linguistic cortical regions were more active late in a trial.

These three explanations are not mutually exclusive. What these findings and their explanations demonstrate is that the effect size matters depending on constraints placed on the cognitive processes.

Language statistics

The effect sizes found for embodied cognition studies are large, warranting the conclusion that cognition is embodied, but this conclusion should be drawn with caution. Different constraints modulate these effect sizes, and throughout this dissertation, I have made the argument that other factors such as language statistics explain cognitive processes. The question needs to be answered whether the effect sizes for language statistics are as large as those for perceptual simulation.

Sixteen studies were extracted on the basis of the following criteria: (1) studies had to be peer-reviewed in a journal or conference proceedings article; (2) include response time experiments; theoretical papers were not included; (3) include language stimuli, where word frequencies were used as an independent variable; (4) include statistically analyzable and comparable results such as ANOVAs or mixed effects models, and their details, such as *F* values, *df*'s, *M* and *SD*. The selected studies are presented in (Appendix 2). These studies included a total of 58 experiments. The average η^2 of the language statistics variables in these studies was large (M = .169, SD = .188). Partial eta squared ($\eta_p{}^2$) for language statistics analyses did not statistically differ from the effect sizes found in the embodied cognition analyses reported earlier, $F(1, 182)$ = .672, p = .280, η^2 = .004, M = .185, SD = .177 vs. M = .169, SD = .188.

As in the perceptual studies, the use of single words again explained the dependent variable of effect size, $F(4, 52)$ = 4.34, p = .042, η^2 = .257, (M = .022, SD = .030 vs. M = .186, SD = .195). Type of response (e.g., RT) $F(4, 52)$ = .005, p = .942, η^2 = .013, (M = .245, SD = .186 vs. M = .158, SD = .191), stimuli configuration $F(4, 52)$ = .584, p = .448, η^2 = .071, (M = .189, SD = .202 vs. M = .136, SD = .175) and method of response (e.g., button press) $F(4, 52)$ = 1.952, *p* = .168, η^2 = .072, (*M* = .446, *SD* = .031 vs. *M* = .156, *SD* = .186) did not explain the effect sizes.

However, as pointed out earlier, η^2 is not ideal for computing effect sizes from studies with different error terms (Roberts & Monaco, 2006). Particularly, because the majority of the language statistics studies used mixed effects models with – consequently – larger degrees of freedom than ANOVAs, η^2 as the measure of effect size underestimates the effect sizes of these studies. Hedge's *g* was again calculated to allow for comparability of the results between studies.

As with the η^2 as the effect size, the average Hedge's *g* was high (*M* = 1.31, *SD* = .805), and so was the correlation between the η^2 and Hedge's *g*, *r* (58) = .79. The same analysis for Hedge's g resulted in significant differences for linguistic and perceptual effect sizes, *F*(1, 175) = 8.07, *p* = .005, η^2 = .044, with embodied factors having weaker effects than linguistic factors, *M* = .965, *SD* = .711 vs. *M* = 1.31, *SD* = .805. This finding is in line with Hutchinson and Louwerse (2013a) and Louwerse (2008) who found that when perceptual simulation and language statistics are compared ceteris paribus, language statistics turns out to be the strongest predictor of cognitive processing.

General discussion

Language processing activates the simulation of perceptual experiences. This conclusion has been drawn in many studies that have argued that cognition is embodied. With the wealth of evidence in favor of perceptual simulations, there can be little doubt that embodiment plays a role in numerous cognitive processes. Questions that have been much less closely investigated include whether cognitive processes always rely on perceptual processes, whether these processes also rely on other factors and to what extent and whether any effects are modulated by individual differences, the nature of the cognitive task, the nature of the stimulus and the time course of processing. This chapter explores these questions.

In line with the Symbol Interdependency Hypothesis, which argues that language has encoded perceptual simulations, language users can rely on language statistics, on perceptual simulation or on both factors in conceptual processing. Indeed, for various studies that have reported an effect of perceptual simulation, a complementary factor, language statistics, explained results equally well or better (Louwerse, 2011a). Because linguistic statistical frequencies are built on perceptual information, with very limited symbol grounding, language users can bootstrap meaning from these statistics, at least when forming quick, good-enough representations.

In the current chapter, I evaluated patterns in the effect sizes of reported studies. For a total of 51 studies including 126 experiments, effect sizes were computed for a perceptual simulation variable. There have been concerns that effect sizes for some embodied cognition studies are small (e.g., Wilson & Golonka, 2013) but the overview shows that effect sizes reported for embodied cognition studies overall tend to be large. However, the conclusion that large effect sizes demonstrate that perceptual simulation therefore explains cognitive processing needs to be put in perspective. A factor that many of our studies have shown to be complementary to perceptual simulation, language statistics, has also been shown to have large effect sizes.

The effect sizes also allow for some exploratory analyses on their nature. In a regression analysis, when single words are used as stimuli, the effect sizes for perceptual simulation are reduced. This is in line with studies that have argued that perceptual simulation is relatively slower than statistical linguistic results (Hutchinson & Louwerse, 2012; Louwerse & Connell, 2011) and dominates in deeper cognitive tasks (Louwerse & Jeuniaux, 2010). Moreover, when the effect sizes of perceptual simulation and language statistics from published studies are compared, the effect sizes for language statistics are significantly higher than for perceptual simulation.

It is important to emphasize that perceptual simulation does still play a role early in processing, and perceptual experiences are still activated in shallow cognitive tasks. Here, the effect of perceptual simulations in comparison to language statistics is less in shallow cognitive tasks that yield faster processing, compared to deeper cognitive tasks that yield slower processing (in which cases language statistics play a less prominent role).

The cognitive sciences can be characterized by debates in which the proverbial pendulum swings from one extreme view to the other. In the late

1990s, the pendulum started to move towards embodied accounts of cognition. But it is necessary to caution against extreme views. Barsalou (1999) took a more moderate stance. He argues three basic approaches to knowledge can be distinguished: classic representational approaches based on amodal symbols, statistical and dynamical approaches and embodied approaches. But as is evident from this chapter, theories should try “to integrate the positive contributions of all three approaches” (Barsalou, 1999, p. 652).

Appendix 1. Effect sizes for 126 embodied cognition experiments

Publication	Experiment	η^2	Hedge's *g*
Beilock and Goldin-Meadow (2010)	1[a]	0.174	0.856
Beilock and Holt (2007)	1	0.094	0.648
	2	0.109	0.671
Bergen et al. (2007)	1[abd]	0.074	0.55
	2[abd]	0.09	0.597
	3[abd]	0.007	0.163
	4[abd]	0.002	0.09
	5[abd]	0.013	0.217
Borghi, Glenberg, and Kaschak (2004)	1[ab]	0.281	1.164
	2[abc]	0.638	2.289
	3[ab]	0.813	2.496
Borreggine and Kaschak (2006)	1[ab]	0.152	0.797
	4[ab]	0.034	0.351
Buccino et al. (2005)	1	0.455	1.487
	2[ab]	0.348	1.327
Connell, and Lynott (2009)	1[abc]	0.056	0.461
Connell, Lynott, and Dreyer (2012)	1[ac]	0.005	0.918
	2[ac]	0.006	1.005
Dijkstra et al. (2004)	1[ab]	0.481	1.691
Estes et al. (2008)	1[abcd]	0.703	2.846
	2[abcd]	0.078	0.562
	3[abcd]	0.465	1.689
Glenberg et al. (2009)	1[ab]	0.133	0.757
	2[ab]	0.304	1.216
	3[ab]	0.06	0.467
	4[ab]	0.112	0.687
Glenberg, Robertson, Jansen, and Johnson-Glenberg (1999)	1	0.409	1.646
	2[ab]	0.056	0.46
	3[ab]	0.378	1.487
Havas et al. (2007)	1[ab]	0.044	0.425
	2[ab]	0.169	0.555
	3a[abc]	0.019	0.271
Holt and Beilock (2006)	1[ab]	0.065	0.52
	2[ab]	0.059	0.506
IJzerman and Semin (2010)	1	0.081	0.574
	2	0.076	0.56
	3	0.062	0.503
	4	0.151	0.803
Jostmann, Lakens, and Schubert (2009)	1	0.111	0.683
	2	0.081	0.577
	3	0.093	0.616
	4	0.176	0.882
Kaschak et al. (2005)	1[ab]	0.097	0.611
	2[ab]	0.109	0.644
Kaschak, Zwaan, Aveyard, and Yaxley (2006)	1[ab]	0.135	0.729
	2[ab]	0.098	0.638
	3[ab]	0.048	0.389
Kaup, Lüdtke, and Maienborn (2010)	1[ab]	0.565	1.999
	2[ab]	0.238	0.965
Kaup, Yaxley, Madden, Zwaan, and Lüdtke (2007)	1[ab]	0.193	0.91

Publication	Experiment	η^2	Hedge's *g*
	2[ab]	0.101	0.58
Koch, Holland, Hengstler, and van Knippenberg (2009)	1[ac]	0.171	0.879
Louwerse and Jeuniaux (2010)	1a[bc]	0.001	1.323
	1a[abc]	0.000	0.062
	1b[bc]	0.353	2.734
	1b[abc]	0.124	1.290
	2a[b]	0.027	1.575
	2a[ab]	1.010	1.196
	2b[bc]	0.031	2.218
	2b[abc]	0.031	2.145
Markman and Brendl (2005)	1[ab]	0.17	0.845
	2[ab]	0.286	0.845
Matlock (2004)	1[ab]	0.409	1.632
Meier and Robinson (2004)	1[ab]	0.134	0.869
	2[ab]	0.26	1.246
	3[ab]	0.118	0.753
	4[abc]	0.108	0.674
Meier et al. (2007)	1[abcd]	0.156	0.828
	2[ab]	0.23	1.042
Meier et al. (2004)	1[abc]	0.854	4.506
	2[abcd]	0.422	1.664
	3[bd]	0.128	0.737
	4[d]	0.257	1.12
	5a[bd]	0.441	1.742
Meier et al. (2004)	1a[abc]	0.357	1.252
	1b[abc]	0.212	0.897
	2a[bc]	0.362	1.32
	2b[bc]	0.327	1.123
	3a[bc]	0.181	0.837
	3b[bc]	0.228	0.921
	4a[abc]	0.029	0.299
	5a[abc]	0.015	0.219
Meteyard, Bahrami, and Vigliocco (2007)	1[abc]	0.292	1.2
Meteyard et al. (2008)	1[ab]	0.03	0.622
	2-4[ab]	0.079	0.924
Myung et al. (2006)	1[ab]	0.129	0.741
Nuthmann and Van Der Meer (2005)	1[c]	0.317	1.305
Pecher et al. (2010)	1[abcd]	0.021	0.286
Pecher, van Dantzig, Zwaan, and Zeelenberg (2009)	1[b]	0.074	0.553
Rapp and Horton (2003)	1[ab]	0.241	1.078
	2[ab]	0.198	0.947
Richardson and Matlock (2007)	1[a]	0.174	0.897
Richardson, Spivey, Barsalou, and McRae (2003)	1[abd]	0.07	0.538
	2[abd]	0.063	0.513
Richter and Zwaan (2009)	1[abc]	0.525	2.01
	2[abc]	0.214	1.011
Santana and de Vega (2011)	1[ab]	0.284	1.171
	2[ab]	0.446	1.73
	3[ab]	0.136	0.762
Schubert (2005)	2[abcd]	0.16	0.8
	3[abcd]	0.263	1.151
	4[abcd]	0.112	0.68
	5[abcd]	0.197	0.953
	6[bd]	0.048	0.442

Publication	Experiment	η²	Hedge's *g*
Sell and Kaschak (2010)	1[ab]	0.074	0.396
	2[ab]	0.031	0.35
Šetic and Domijan (2007)	1[abcd]	0.27	1.156
	2[abcd]	0.307	1.294
Stanfield and Zwaan (2001)	1[ab]	0.15	0.782
Taylor, Lev-Ari, and Zwaan (2008)	1[a]	0.051	0.457
van Dantzig et al. (2008)	1[a]	0.068	.514
van Dantzig, Zeelenberg, and Pecher (2009)	1[ac]	0.069	.517
	2[ac]	0.071	.526
	3[ac]	0.145	.766
Vermeulen, Mermillod, Godefroid, and Corneille (2009)	1[abc]	0.868	4.791
Yaxley and Zwaan (2006)	1[ab]	0.135	0.763
	1[abcd]	0.145	0.783
	3[abc]	0.089	0.575
Zwaan, Madden, Yaxley, and Aveyard (2004)	1[ab]	0.081	0.572
Zwaan et al. (2002)	1[ab]	0.25	1.079
	2[a]	0.076	0.531
Zwaan and Taylor (2006)	1[a]	0.221	1
	2[a]	0.071	0.536
	3[ab]	0.188	0.93
	4[ab]	0.048	0.427
	5[ab]	0.046	0.418

[a]Response time; [b]Button press; [c]Word; [d]Vertical configuration.

Appendix 2. Effect sizes for 58 language statistics experiments

Publication	Experiment	η²	Hedge's *g*
Connell and Lynott (2013)	1	0.014	1.096
	2	0.013	1.033
Estes (2003)	1	0.153	0.781
	2	0.274	1.162
Gagné (2002)	1	0.238	1.431
	2	0.308	1.814
Hutchinson and Louwerse (2012)	1	0.454	2.821
	1	0.406	2.286
	2	0.524	3.257
	2	0.402	2.408
Hutchinson and Louwerse (2013a)	1a	0.41	2.568
	1b	0.489	2.814
	2a	0.417	1.752
	2b	0.265	1.28
	3a	0.485	1.714
	3b	0.478	1.936
	4a	0.369	1.867
	4b	0.439	2.167
Hutchinson and Louwerse (2013b)	1	~ 0	0.026
	1	0.001	0.53
	1	~ 0	0.111
	1	0.004	0.97

Publication	Experiment	η^2	Hedge's *g*
	1	0.001	0.47
	1	0.002	0.492
	2	~ 0	0.456
	2	0.009	1.268
	2	0.003	1.131
	2	0.02	1.924
	2	0.004	0.796
	3	0.002	0.506
Hutchinson and Louwerse (2013c)	1	0.001	NA
	1	0.001	NA
	1	0	NA
	1	0.005	NA
	1	0.004	NA
	1	0.002	NA
Louwerse and Hutchinson (2012)	1	0.005	1.007
Louwerse and Jeuniaux (2010)	1a[1]	0.146	1.559
	1a[2]	0.152	1.543
	1b[1]	0.099	1.124
	1b[2]	0.112	1.225
	2a[1]	0.190	0.324
	2a[2]	0.009	1.196
	2b[1]	0.086	0.938
	2b[2]	0.207	2.145
Louwerse (2011b)	1	0.203	NA
Santiago, Lupáñez, Perez, and Funes (2007)	1	0.272	1.172
	1	0.048	0.828
	2	0.58	2.936
Solomon and Barsalou (2004)	1	0.273	2.466
Tagalakis and Keane (2006)	1	0.325	1.273
	2	0.019	0.245
Tillman et al. (2013)	1	0	0.196
	1	0.006	0.728
	1	0.001	0.318
	1	0.004	0.573
Wu and Barsalou (2009)	1	0.425	1.071
	2	0.468	1.257

[1] = error rate; [2] = response times.

Chapter Six

Linear Mixed Models

Abstract

The language-as-fixed fallacy is the failure to include items as a random factor in statistical analyses, yielding unsuccessful generalization past those items in an experiment. The problem was outlined five decades ago, and different solutions to this problem have been applied by researchers in the psychological sciences, such as separate by-subject (F_1) and by-item (F_2) analyses and combined *MinF'* analyses. This chapter aims to shed light on the conditions under which results that are obviously significant for a linear mixed model might beget insignificant results for F_1 and F_2 analyses, and vice versa, by manipulating the effect of treatment in a variety of simulated datasets. The second aim is to estimate the number of publications in the current literature that might be reporting incorrect results simply from using an F_1 and F_2 analysis. Based on simulations of datasets, a regression formula was estimated that allows for predicting the significance of a linear mixed model from F_1 and F_2 values. Results suggested that approximately 34% of the studies using F_1 and F_2 analyses might be subject to a Type I error, with an unknown number of unpublished studies being subject to a Type II error.

This chapter is based on:

Hutchinson, S., & Louwerse, M. M. (2015). *Publish or Perish: Consequences of Considering Sampling Errors.* Manuscript submitted for publication.

Hutchinson, S., Wei, L., & Louwerse, M. M. (2014). Avoiding the language-as-a-fixed-effect fallacy: How to estimate outcomes of linear mixed models. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Introduction

In the previous chapter I clarified the conditions under which perceptual simulations are most important, showing that effect sizes of language statistic factors tend to be as large or larger than those of perceptual stimulation factors. Continuing in the vein of statistics, this chapter also examines results from a number of embodied cognition studies by explaining and justifying the methodological choices that I made in the research presented here, arguing that linear mixed models provide the most suitable analytical approach to provide answers to the questions posed in this manuscript. This chapter is not intended to address varying random effect structures, or random slopes by treatment, but instead focuses on very simple models. In this final quantitative chapter, I will focus on presenting several statistical simulations and argue that the analyses used in this manuscript provide more accurate and reliable results than the standard models used in the literature.

Considering the large number of investigations in the literature that have used a test that ignored the sampling error of the materials, the laws of probability ensure that a percentage of the significant results were due to such chance variance. Or in other words, a percentage of these results could not be replicated using a different sample of language materials (Coleman, 1964, p. 226). Coleman (1964) recognized that researchers in the psychological sciences properly specified participants as random factors in their analyses, yet variance in experimental items (e.g., word stimuli, sentence stimuli, picture stimuli, etc.)

was all but ignored. The consequence was that experimental results were published for which the statistical results did not correctly account for all random error in the data, potentially leading to erroneous conclusions in the literature. The failure to indicate items as a random factor, yielding unsuccessful generalization past specific items included in a particular experiment, is known as the "language-as-fixed-effect fallacy" (Clark, 1973). Coleman (1964) and Clark (1973) argued that both participants and items should be treated as random factors. Just as participants in an experiment do not represent an entire population, items in an experiment are by no means representative of the entire population of possible stimuli (Baayen, Davison, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013). Clark (1973) proposed a simple solution to the language-as-fixed-effect fallacy, recommending a calculation of a mixed F value that included a model with a random participant factor (F_1) and one with a random item factor (F_2). This estimate of a combined F value is referred to as *minF'*.

Since the pleas by Coleman and Clark, conclusions from much experimental work in cognitive science, especially regarding embodied cognition has relied on the use of separate subject and item (F_1 and F_2) regression analyses. Even though this seems to solve the language-as-fixed fallacy, the solution of reporting F_1 and F_2 analyses separately is rather surprising, considering Clark (1973) never proposed F_1 and F_2 analyses should be reported separately. Instead, he argued for the use of a combined *minF'* score. The fact that F_1 and F_2 were intended as intermediate steps used to calculate *minF'* and not as a replacement for *minF'* is what researchers seem to have neglected, perhaps because *minF'* is often (erroneously) thought of as an overly conservative estimate of the combined F value it approximates (Raaijmakers, Schrijnemakers, & Gremmen, 1999). Moreover, when separate F_1 and F_2 values are reported, ambiguous situations can emerge in which F_1 reaches the specified alpha level (e.g., $p < .05$), but F_2 does not, or vice versa (Locker, Hoffman, and Bovaird, 2007), making it difficult to draw meaningful conclusions. It is not uncommon to see emphasis placed on those F values that are significant, while insignificant F values tend to be ignored. In short, whereas Coleman (1964) cautioned against the language-as-fixed-effect fallacy, and Clark (1973) reiterated this warning and proposed a solution with *minF'*, experimental studies tend to only report F_1 and F_2 values - two components used to calculate *minF'*.

In the meantime, a solution to the separate F_1 and F_2 values and to the language-as-a-fixed-effect fallacy has been readily available in linear mixed models. Linear mixed models (LMM), first having seen widespread use in in biomedical research in the 1950s, are also known as multilevel models, hierarchical linear models, mixed effects models, or variance component models and have recently been advocated in the psychological sciences (Baayen, Davidson, & Bates, 2008; Brysbaert, 2007; Locker, Hoffman, & Bovaird, 2007; Pinherio & Bates, 2000; Richter, 2006).

With different measures such as F_1, F_2, *minF'* and LMM available to address sampling errors in experimental studies, it remains unclear what the impact is on the psychological sciences. Conclusions drawn from significant results should in hindsight not have been published, because their findings should have been attributed to chance and — equally importantly — unpublished studies should in hindsight have been published because their findings did reach the specified alpha level and should be considered significant. This chapter aims to give insight in the probability of significant results in one type of analysis (F_1, F_2, and *minF'*, or LMM), based on the findings from another analysis in order understand under what conditions results that are significant for one model might beget significant results for other models.

It is critical to point out here that I do not want to emphasize the importance of the set alpha level. Reported effects extensively rely on p values in order to quantify the relevancy of an effect (Cohen, 1994). Convention dictates that p values are deemed significant when $p < .05$ and are sometimes considered marginally significant at p values of $p < .10$. Given this tradition that the alpha level is set at $p = .05$, and given that findings that exceed this threshold are typically considered not significant and oftentimes therefore not publishable, this chapter will discuss the notion of significant and insignificant results.

Insights in estimates of the results when the analyses had included both participants and items as random factors would provide researchers in psychological sciences with a rule-of-thumb whether their conclusions would hold if their data were properly reanalyzed. Moreover, such a rule-of-thumb allows for measuring the impact of different statistical analyses on the field of psychological sciences. Before introducing this rule-of-thumb and investigating the impact of the effect of different statistical methods on studies in the psychological sciences, I will first turn to a more detailed description of the different measures addressing sampling error.

minF'

MinF' is an *F* value that is calculated from the familiar F_1 and F_2 values and is computed as $(F_1 * F_2) / (F_1 + F_2)$, where F_1 is the *F* value of the by-participant ANOVA analysis and F_2 is the *F* value of the by-item ANOVA analysis. However, the *minF'* value suggested by Clark (1973) is only an approximation of another value, *F'*. *F'* is derived from the formula $(MS_T + MS_{S*I*T}) + (MS_{T*S} + MS_{I*T})$, whereby MS_T is the mean square of the treatment effect, MS_S is the error term of the participants, and MS_I is the error term of the items. However, *F'* is difficult to calculate when dealing with a large dataset or missing data (Raaijmakers, Schrijnemakers, & Gremmen, 1999) which is why *minF'* is calculated instead.

What makes this even more complex, is the fact that *F'* is also an approximation of a combined *F* value, and like *F'*, the combined *F* it approximates is also difficult to compute when data are missing. *minF'* is an approximation of an approximate value, and is a conservative (minimum lower bound) approximation of *F'*. *F'*, in turn, is also a conservative approximation of the combined *F* it approximates. Therefore, the significance for *minF'* values must be calculated independently from F_1 and F_2 values. Simply put, *minF'* does not automatically inherit significance just because F_1 and F_2 are independently significant.

F_1 and F_2

In much of the psychological science literature, particularly in the field of psycholinguistics, it has become convention, to report only the seemingly less conservative (and therefore more often significant) F_1 and F_2 values instead of the combined *minF'* values, despite the fact that the chances of making a Type I error increase. There are two reasons that may explain this practice of reporting F_1 and F_2. First, as speculated by Raaijmakers, Schrijnemakers, and Gremmen (1999), there might be a lack of understanding on the part of the researcher, in that researchers simply may have misunderstood that they are supposed to report *minF'* and not only the components used to calculate *minF'*. Second, and equally problematic, is the fact that researchers regard *minF'* as too conservative and rather than reporting an insignificant *minF'* value, they rather report significant F_1 and F_2 values, or worse, a single significant F_1 or F_2 value.

The correct *minF'* value since introduced has steadily declined in use. Nowadays it is virtually unseen in published articles, being replaced by the separate F_1 and F_2 analyses. Raaijmakers, et al. (1999) reported that out of 220 articles from 1993-1997 in the *Journal of Memory and Language* that mention F_1 and F_2 a total of 120 papers report F_1 and F_2 values, ignoring *minF'* altogether.

Linear mixed models

Linear mixed models are more powerful than linear regressions because they solve the language-as-fixed fallacy by allowing researchers to consider both participant and item error simultaneously in one model. In doing so, researchers increase model fit by driving down random error. In essence, linear mixed models do not treat items (or participants) as a fixed effect, thereby offering an alternative to the infrequently used *minF'*.

In addition to offering a solution to the language-as-fixed-effect fallacy, these models also have several additional advantages compared to traditional models. First, linear mixed models can accommodate more complicated nested and crossed designs (Quené & van den Bergh, 2008). Furthermore, linear mixed models allow for missing data at random and do not need to perform listwise deletion. They can be further extended to allow for time-varying covariates and they accurately present the relationships between variables over time. Mixed models also easily allow for clustering, longitudinal or repeated measures as well as specific covariate structures. Finally, they are able to generalize non-normal data and do not assume independent observations, thereby being more applicable to a wide range of datasets.

Recent work by Baayen, Davidson, and Bates (2008) encouraged researchers to utilize linear mixed models by clearly demonstrating the differences in the results of different models that were applied to the same datasets. West, Welch, and Gałecki (2006) take a similar approach by demonstrating how to run linear mixed models in various software packages. But despite software and tutorials for running linear mixed models being readily (and sometimes freely) available in *R*, *SPSS*, *SAS, MLwiN* and other packages, and despite the convincing arguments favoring the benefits of linear mixed models (Baayen, 2008a; Brysbaert, 2007; Richter, 2006; West, Welch, & Gałecki, 2006; Winter, 2013), the use of mixed models is still not widespread. In fact, nearly fifteen years after Raaijmakers, Schrijnemakers, and Gremmen (1999) reported that most published articles in the *Journal of Memory and Language* failed to report *minF'* values, in the same journal between 2012-2013, out of 56 published articles mentioning F_1, F_2, *minF*, or mixed models, more than half (n = 30) still report F_1 and F_2 values, three of which also report *minF'*. At the same time, 26 papers correctly report results from linear mixed models, suggesting that at least some progress is being made with correctly addressing sampling error.

The advice from Coleman (1964), Clark (1973), Raaijmakers, et al. (1999) and Baayen et al. (2008) to researchers still reporting F_1 and F_2 values to correctly reanalyze data with LMM will likely not be received enthusiastically, as it is unclear whether the previous conclusions drawn from the results would in fact still hold. Ideally, it would be desirable to estimate, on the basis of F_1 and F_2 values, whether mixed effect models would generate significant results and vice versa. Such an estimate would clearly not replace a reanalysis of the data with mixed models, but could serve as an estimate of the effect of a proper statistical method on the findings, advising researchers and allowing them a measure of the impact of F_1 and F_2 analyses has had on the field of psychological science. This might subsequently motivate a mixed model analysis of the original data, or a replication of the experiment with new data using the proper statistical model.

This chapter has two goals. First, by manipulating the effect of treatment in a variety of simulated datasets, it aims to shed light on the conditions under which results that are obviously significant for a linear mixed model might beget insignificant results for F_1 and F_2 analyses, and vice versa. Second, this chapter estimates the number of publications in the current literature that might be reporting incorrect results simply from using an F_1 and F_2 analysis. It is well discussed that in the cognitive science literature, researchers extensively rely on *p*-values in order to quantify if an effect is relevant (Cohen, 1994; Wagenmakers, 2007). In this chapter, I calculate *F* and *p* values from simulated datasets using a linear mixed model and using more traditional analyses to generate formulas for estimation (F_1 and F_2, *minF'*). These formulas are then used to determine if results from linear mixed models can indeed be estimated from traditional analyses. Next, I estimated how many results of articles published in *Psychological Science* between 2004-2014 might be impacted from using linear mixed models.

Simulations

The first step was to simulate datasets in order to calculate formulas from which mixed model *F* and *p* values can be estimated. Simulations took place using three experimental designs that imitated semantic judgments of words in which response times from participants served as dependent variable. If *p* and *F* values can be estimated across models, then they are most likely best estimated with a simple model and, following the principle of parsimony, therefore a design was selected with only one independent variable and one dependent variable. Two within-participant designs and one between-participant design were included in order to remain consistent with the most common experimental designs (Howitt & Cramer, 2011). To match standard designs in the current literature, for within-participant (n = 20) and between-participant designs (n = 40) different items were used in each treatment condition. The second within-participant design (n = 10) had the same items in each treatment condition. The number of participants for each condition ranged from 10 – 40, with 40 participants in the between participants design and 10 and 20 participants in the two within subject designs. This was to ensure that the number of participants in each condition resulted in the same number of data points for all designs.

Data for each of the three designs above was simulated 100 times to generate 300 simulated datasets (100 simulations * 3 different designs = 300 total simulations). Data was also simulated for each of the three designs with normally distributed errors (an ideal scenario for parametric tests), positively skewed errors, negatively skewed errors, and random errors (that is, 100 simulations * 3 different designs * 4 error distributions = 1,200 total simulations). In addition, to these complete datasets, simulations were run for data with missing values of up to 10% (100 simulations * 3 different designs * 4 error distributions * 2 missing data scenarios = 2,400 total simulations) (Finch, 2010). Missing data and non-normal errors were included in order to increase the generalizability of the simulations, and their applicability to the real word, as imperfect distributions are sometimes found in real datasets. Finally, these 2,400 simulations were simulated ten times (2,400 simulations * 10 effect of treatment manipulations = 26,400 total simulations), because manipulating the effect of treatment (E_T) from no effect ($p > .10$) to a strong effect ($p < .01$) required 10 iterations.

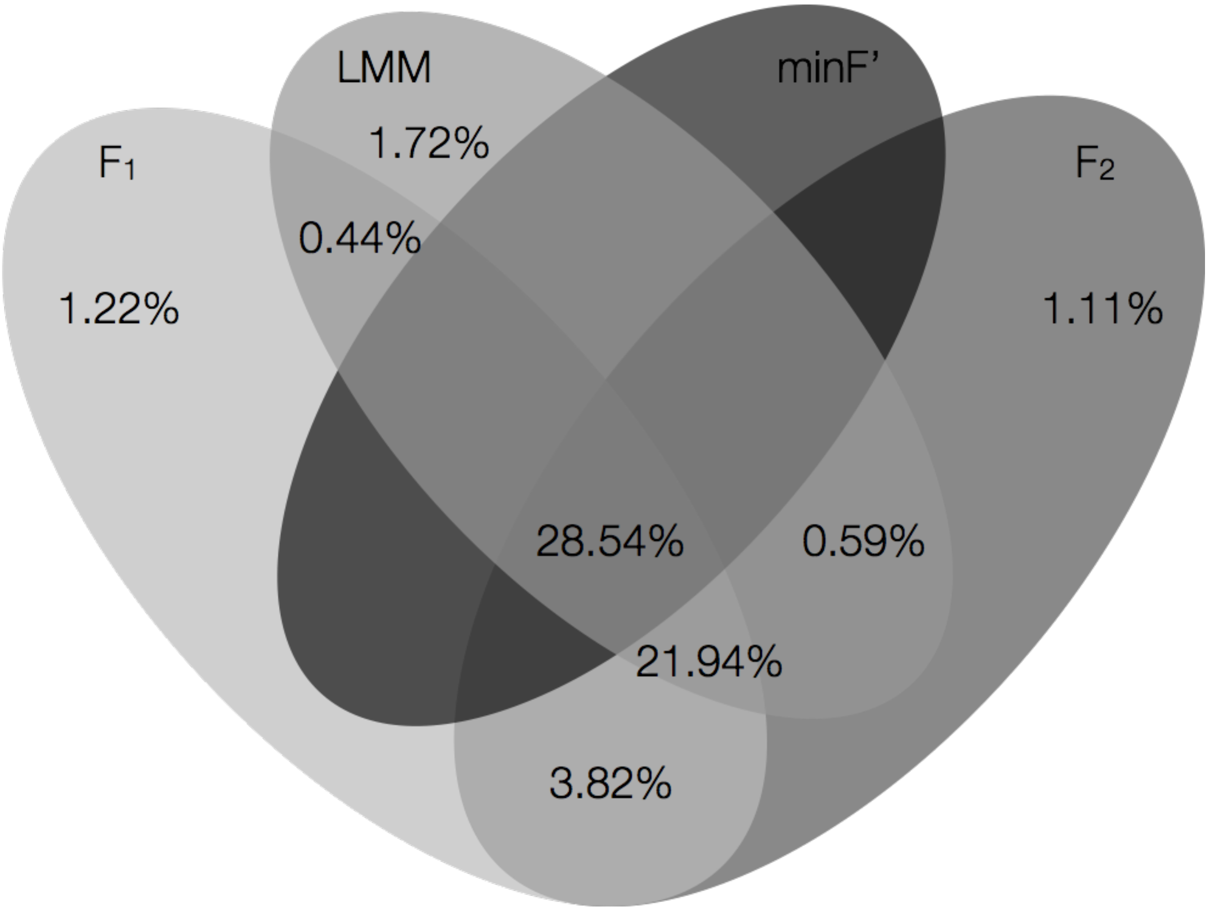


Figure 1. Percentage of significant cases out of the total 14784 significant simulations across F_1, F_2, *LMM*, and *minF'* models. The percentages in the grey areas represent shared significance between models. Overall, F_1 and F_2 were significant 56% of the time, LMM were significant 53% of the time, and minF' was only significant 29% of the time.

Each dataset had the same initial values from which to generate a simulated dependent variable. A linear model has the following structure: $Y = Y_0 + E_T + E_S + E_I + E$, where Y is the dependent variable (e.g., response time), Y_0 is the expected mean response time with no treatment, E_T is the strength of the effect, E_S is the by-participant error, E_I is the by-item error, and E is the by-observation error. Y_0 was set to 400 ms for each response, a recognized response latency for semantic judgments of words (Kutas & Hillyard, 1980). All normally distributed errors (by-participant error (E_S), by-item error (E_I), and by-observation error (E)) were set to be randomly generated numbers centered at 0, where the *SD* of the error was a random number ranging between 0 and 20. Recall the distribution of the errors was determined in advance to be either normal, positively skewed, negatively skewed, or random. The strength of the effect (E_T) was manipulated so that each design was used 10 times, ranging from no effect of the independent variable, to all 100 cases resulting in highly significant effects at $p < .01$.

F_1, F_2, *minF'*, and linear mixed models were all computed using lme4 in the languageR package (Bates & Sarkarin, 2007). Significance was estimated for both the two tailed MCMC probability as calculated from the pvals.fnc function in the languageR package (Baayen, 2008b) and for the t-test using Satterthwaite's approximation for the degrees of freedom found in the lmerTest package (Kuznetsova et al., 2013). Both methods generated the same results.

Four models, an F_1 model, an F_2 model, a *minF'* model and a LMM were conducted on the data for each of the simulations. The number of significant cases out of the 26,400 total simulations for each model is represented in Figure 1. As Figure 1 shows, *minF'* values were the most conservative, with significance at $p < .05$ in 28.54% of the data. In fact, in these simulations, *minF'* values were never significant when the other measures were also not significant. Such findings are in line with the fact that *minF'* has been argued to have reduced power compared to all other models (Wickens & Keppel, 1983). F_1 and F_2 models were least conservative, resulting in significant *p* values of $p < .05$ for 55.97% of the time for F_1 models and for 56.01% of the time for F_2 models. Results from both F_1 and F_2 models had a large overlap, with each only being independently significant from the other less than 1% of the time. Linear mixed models were equally conservative, with significant *p* values set at $p < .05$ being found in 53.23% of the cases out of 26,400 simulations. A large overlap was found between linear mixed models and F_1 and F_2 models. Linear mixed models were independently significant less than 2% of the time. In other words, linear mixed models and F_1/F_2 results were the most similar in terms of identifying significant results.

Although F_1/F_2 models always detected more significant results than did other models, it is important to note that when E_T was small linear mixed models detected more significant effects than did F_1, F_2, or *minF'* (see Figure 2). These findings suggest that findings reported with significant F_1 and F_2 results, are likely significant when data is analyzed with linear mixed models, but this is not the case for *minF'*. These reported results are for all simulated models, with random and skewed data.

Next, I investigated whether the results from linear mixed models could be estimated from the output of the other models. There are several possible factors impacting whether significance can be estimated in one model based on the results from another, including the experimental design, the size of the

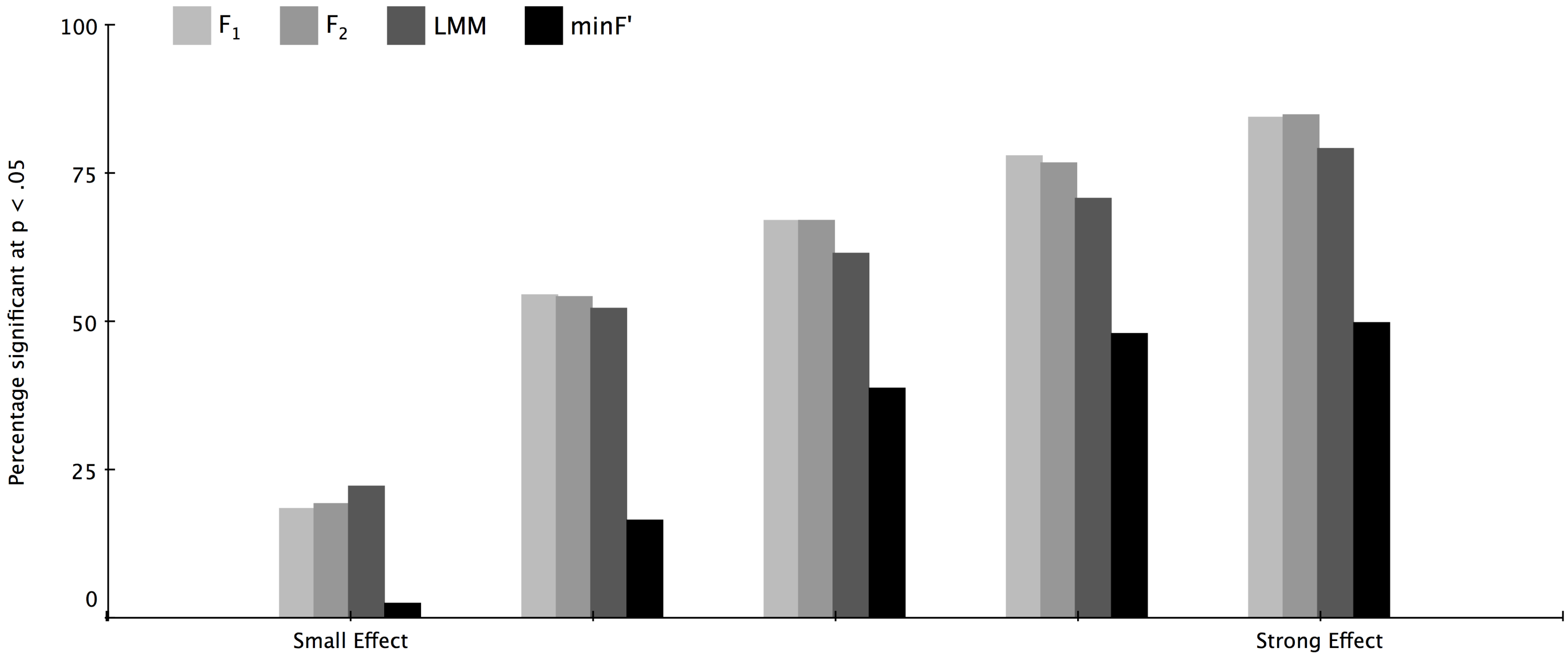


Figure 2. The total percentage of significant p values for each model. The E_T is split into five bins, from a small effect of E_T (first bin) to a strong effect of E_T (last bin). * denotes value is significantly different from other groups at $p < .05$.

effect, the number of factors, and the degrees of freedom. However, I tried to estimate the outcome of linear mixed models (in these simple models) from respectively very little information (i.e., *p* and *F* values).

First, to determine whether F_1, F_2, and *minF' F* values estimated *F* values in linear mixed models, I entered F_1, F_2, and *minF' F* values in a linear regression model. When estimating *F* and *p* values, it is likely that researchers only have one type of *F* value (either F_1/F_2 or *minF'*, not both). Since F_1/F_2 or *minF'* is normally used to estimate the likely output of a linear mixed model I decided to run analyses where F_1/F_2 and *minF'* were entered into separate analyses. F_1 and F_2 models significantly estimated mixed effect *F* values for F_1, $F(1, 26397) = 2622.46$, $p < .001$, $\eta^2 = .09$, $r = .30$, and for F_2, $F(1, 26397) = 2341.59$, $p < .001$, $\eta^2 = .08$, $r = .28$, resulting in Equation 1. Effect sizes (both η^2 and *r* reported here) were moderate.

$$F = (2.8494 * F_1) + (0.5273 * F_2) - 4.2842$$

To see if the same factors were able to estimate significance, as degrees of freedom were calculated differently in linear mixed models than for standard regressions, all 26,400 *p* values were entered into a regression model and found that F_1 and F_2 *p* values significantly estimated LMM *p* values in linear mixed models for F_1 *p* values, $F(1, 26397) = 2622.46$, $p < .001$, $\eta^2 = .09$, $r = .30$, and for F_2 *p* values, $F(1, 26397) = 2341.59$, $p < .001$, $\eta^2 = .08$, $r = .28$, resulting from Equation 2.

$$p = (0.51654 * F_{1p}) + (0.48786 * F_{2p}) + 0.01324.$$

For *minF'*, F values were also significantly predicted, F (1, 26397) = 303986.8, $p < .001$, $\eta^2 = .92$, $r = .96$, as were p values, F (1, 26397) = 43827.42, $p < .001$, $\eta^2 = .62$, $r = .79$, resulting in Equations 3 and 4.

$$F = (6.54332 * minF') - 3.49908$$

$$p = (0.62379 * minF') - 0.08438.$$

Entering F_1/F_2 and *minF'* together in the model also resulted in significant predictions of F values (for F_1 $F(1, 26396) = 1824.14$, $p < .001$, $\eta^2 = .06$, $r = .25$, F_2 F (1, 26396) = 599.27, $p < .001$, $\eta^2 = .02$, $r = .14$, and *minF'* F(1, 26396) = 346.70, $p < .001$, $\eta^2 = .01$, $r = .10$) and p values (for F_1, $F(1, 26396) = 2578.61$, $p < .001$, $\eta^2 = .09$, $r = .30$, F_2, $F(1, 26396) = 2332.89$, $p < .001$, $\eta^2 = .08$, $r = .28$, and *minF'*, $F(1, 26396) = 6692.88$, $p < .001$, $\eta^2 = .20$, $r = .45$).

Despite the fact that many factors contribute to whether or not F and p values can be estimated from other F and p values, including design, the size of the effect, the number of factors, and the degrees of freedom, for simple models, estimates for mixed model results can be made from other models (F_1, F_2, *minF'*).

Estimates

The predictive power of the formulas derived from the simulated data presented in the previous section were next tested on a different dataset, the splitplot dataset previously used by Baayen, Davidson, and Bates (2008) and freely available in the languageR package (Baayen, 2008b). The experimental design for this dataset involved two counterbalanced lists of words, each with 40 words. Each list consisted of related prime words and unrelated prime words. Twenty participants were tested on one list, or the other.

Because this is only one dataset, RT values for the splitplot dataset were simulated 1000 times using the parameters of the original data such that all simulated data were generated from the distribution of the original mean and *SD* for each parameter. The effect of the IV (E_T) was set randomly so that models would vary from a no effect of treatment at $p > .10$ to a strong effect of treatment at $p < .01$. One thousand simulations of linear mixed models predicting RT with the priming condition as a fixed factor and participant and item as random factors were conducted on the splitplot dataset. Regressions were also conducted for actual F_1 and F_2 values.

Values of significance for each dataset were estimated from the previous formulas and compared these values to the actual output from 1,000 simulations of the dataset provided in splitplot. As can be seen from Figure 3, predicting F and p values for linear mixed models from the F_1/F_2 analyses is almost perfect for simple designs with one independent variable, one dependent variable, and normally distributed errors. This demonstrates that the formulas generated from the simulated data above are able to successfully predict F and p values for real world datasets.

Current Literature Analysis

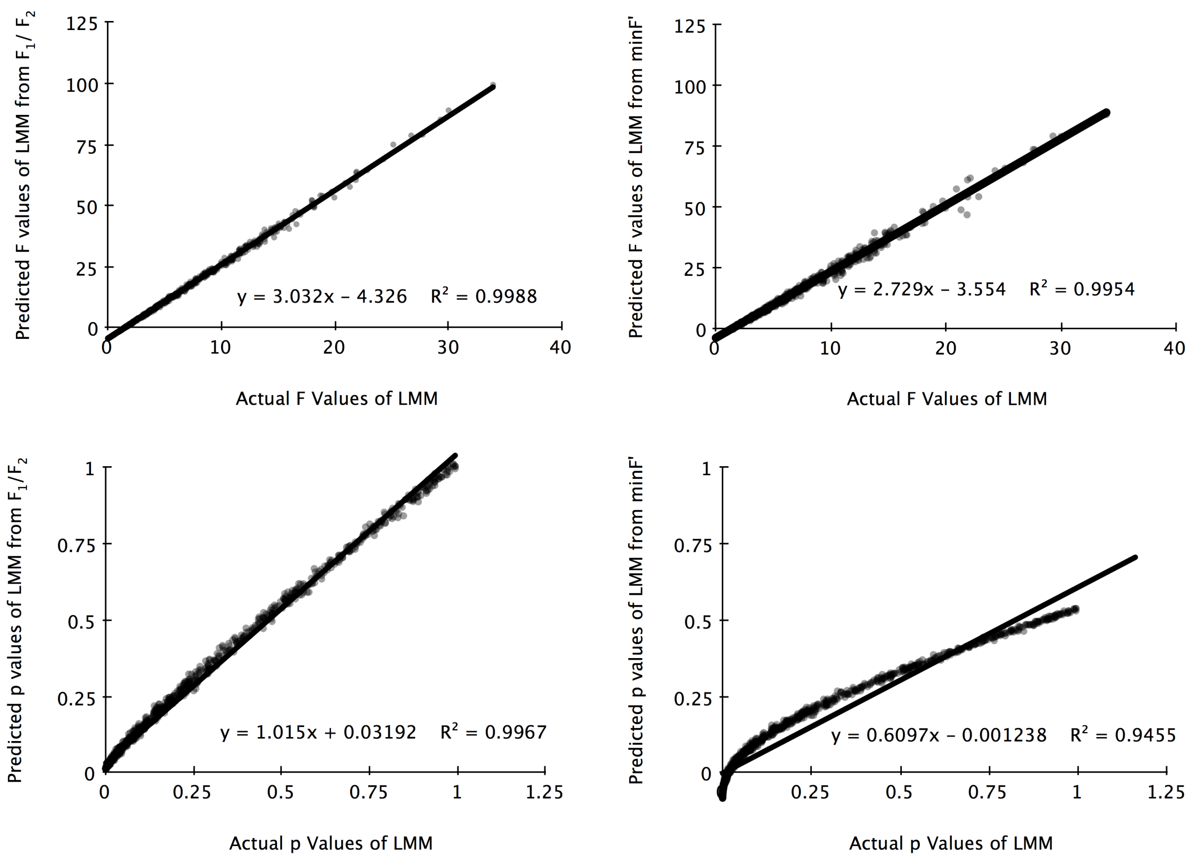
Knowing that significance can be reliably estimated from simulations in the splitplot dataset, next I estimated how many findings in publications in the psychological sciences would be affected by a different statistical analysis. That is, should the results from reported F_1 and F_2 analyses with marginally significant results actually be attributed to Type I errors when considering linear mixed models? And are results reported with highly significant effects ($p < .05$), marginally significant effects ($.10 < p > .05$), or insignificant effects ($p < .10$) equally vulnerable to having dramatically different results when analyzed used linear mixed models?

A total of 2,394 articles published between 2004-2014 in *Psychological Science* were collected. Of the 2,394 papers only 109 papers (5%) reported F_1 and F_2 analyses, and none used *minF'*. Of those 109 papers, only 34% of the articles report all F_1 and F_2 actual values (many report incomplete rounded values, such as $F < 1$). Because these 109 articles often reported multiple experiments, 195 total experiments reported F_1 values and 195 reported F_2 values. Interestingly, of those 195 values, 20% of the pairs of F_1 and F_2 values resulted in ambiguous scenarios in which one F (F_1 / F_2) value was significant, while the other F value (F_2 / F_1) was not, not allowing definitive conclusions to be drawn regarding an effect.

For all 101 pairs of F_1 and F_2 values, F and p values were estimated for a linear mixed model. All significant effects at $p < .01$ remained significant at $p < .01$ when a linear mixed model was used instead of F_1 and F_2 analyses.

For 57 pairs of F_1 and F_2 values significant at $.01 < p < .05$, 11% ($n = 13$) previously reported as significant at $p < .05$ were no longer significant in linear mixed models.

However, when examining the pairs of F_1 and F_2 values where one F (F_1 / F_2) value was significant at $p < .05$ while the other (F_2 / F_1) was not significant, 95% ($n = 108$) of the p values no longer yielded significant results



Figures 3A-3D. F and p values estimated from F_1/F_2 (left) and *minF*' (right) plotted against actual F and p values for the dataset splitplot.

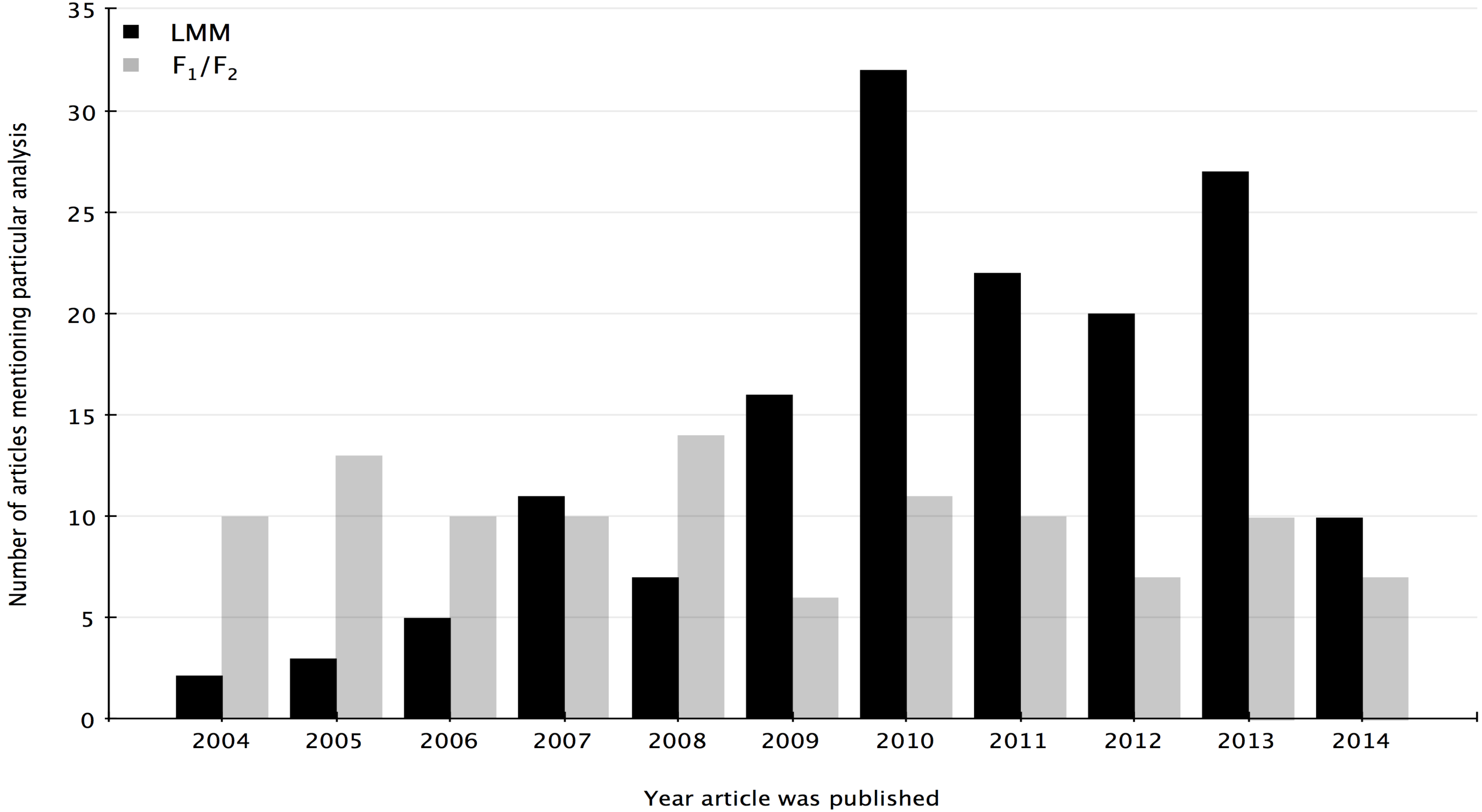


Figure 4. Number of articles published in *Psychological Science* between 2004 - April 2014 that mention linear mixed models or F_1/F_2 analyses.

in linear mixed models.

In conclusion, 5% of the articles published over the last decade in *Psychological Science* include studies potentially incorrectly reporting by-item and by-subject analyses. Even though 5% might seem low, this percentage does not reflect the importance of the issue, but rather the representation of studies in *Psychological Science* for which by-participants and by-item variance applies. For instance, the issue discussed here likely applies to 100% of the psycholinguistic studies.

For the studies that have reported conclusions in *Psychological Science* based on the results of experiments, the validity of the findings in 34% of the cases can be questioned. That is, in a third of the cases in which significant findings have been reported, the suspicion of a Type I error can be raised.

It is unclear how many studies that have not been published might be subject to a Type II error, because we simply have no record of those studies and because there is likely a practical bias to report studies that are marginally significant than to not report them. However, what the prediction of Type I error suggests is that Type II errors should certainly not be ruled out. That is, results that have currently not been published because they were not deemed significant in F_1 and F_2 analyses should perhaps have been published according to a *LMM* analysis.

Conclusion

In this chapter, researchers are recommended to analyze their current data and reanalyze past data that was originally reported as F_1, F_2 or *minF'* using linear mixed models. This suggestion is not only relevant for work on embodied cognition, but for the cognitive science community in general. Obviously such a suggestion might not be eagerly considered, for instance because it is assumed that the results would be similar regardless of the model, or because of the unease that results might not reach the set alpha threshold that was reached in the previously reported analyses. Moreover, studies previously not published because results seemed to be attributed to chance, should perhaps have been reported because the results were in fact significant. On the basis of F_1 and F_2 values and *minF'* values it was therefore estimated whether linear mixed effect models would generate significant results. Indeed not only were *F* values estimated, but also *p* values. These estimates are not intended to replace a reanalysis of the data, but rather they are intended to motivate researchers to analyze and properly reanalyze data using linear mixed models.

The findings show that for simplistic models an *F* or *p* value from a linear mixed model can be accurately estimated from the same values from more traditional analyses. It is important to recognize that this chapter only demonstrates this for the most simple of designs, and that with more complexity, it is likely that it becomes more difficult to accurately estimate *F* and *p* values for linear mixed models. Nevertheless, the strength of the relationship between F_1 / F_2 or $minF'$ and the F value from a linear mixed model is not unexpected, as all of these F values are calculated from the same dataset in a similar way. The same logic stands for the p values. This at least suggests that it might be possible to estimate F and p values of linear mixed models from more complex designs.

This chapter has also elaborated upon some of the benefits of linear mixed models, and suggested its use over alternative traditional methodologies such as F_1 and F_2 analyses. Sometimes F_1 is the proper analysis to use, for instance when items are nested in participants and participants are nested in treatments (Clark, 2008, p. 348), or when items are properly counterbalanced or matched. It is nevertheless important for researchers to understand when particular analyses are appropriate to use and when they are not. Even more practically, linear mixed models provide some benefits to researchers with regard to the flexibility and robust nature of the analysis.

The issues discussed in this chapter impact the current literature. Based on a sample of articles published between 2004-2014 in *Psychological Science* to which F_1 and F_2 apply, the results demonstrated that approximately a third of the reported F_1 and F_2 analyses, when reanalyzed using linear mixed models might result in alternative conclusions.

Reported effects usually rely on p values in order to quantify if an effect is relevant for publication (Cohen, 1994; Wagenmakers, 2007). Despite the problems with null hypothesis significance testing, researchers continue to place emphasis on the significance of p values. Usually p values are deemed significant when $p < .05$ and findings that exceed this threshold are considered marginally or not significant and therefore are much less likely to make it in print. That is, undesirably the consequences of considering sampling errors can be the difference between publish or perish.

Chapter Seven

Conclusion

In the previous chapters I have brought to light some of the complex issues currently relevant to the debate on symbolic versus embodied cognition. In this final chapter I will summarize the findings and conclusions presented in this dissertation with the goal of demonstrating that the question of whether cognition is symbolic or embodied is a moot point. Instead I hope to emphasize that both linguistic and perceptual representations are always employed during language processing, and like in this dissertation, future research should also continue to explore what additional factors impact the relative importance of linguistic or perceptual representations.

I began by asking under what conditions symbolic and embodied accounts of cognition explain how words attain meaning. From the work presented here, it should become clear that mental representations are neither exclusively embodied nor exclusively linguistic but rather that mental representations are both embodied and linguistic, but under different conditions. The specific focus of this dissertation investigated how word meaning is established by both symbolic and embodied accounts working together in concert. In particular, the previous chapters examined how different modulators impact how much we rely on linguistic and embodied representations. In particular, I asked if both linguistic and perceptual representations impacted by 1) the time course of processing 2) the spatial presentation of stimuli 3) individual differences or 4) the orientation of stimuli. I also aimed to determine whether linguistic and perceptual representations are independent processes that work together to establish word meaning.

Chapter 2 illustrated how both linguistic and perceptual factors can explain experimental results. This chapter demonstrated that symbolic representations explain results that are generally attributed to effects of embodied cognition. In three experiments, participants participated in SNARC tasks for numerical stimuli and high and low frequency words. Although the SNARC effect was replicated, providing support for a fundamentally embodied account, for all three experiments, linguistic factors were able to explain subject RTs. These findings not only demonstrated that linguistic symbolic representations can account for human RTs on embodied tasks but also showed that integrated theories of cognition can more easily account for effects together than each theory can on its own.

Chapter 3 addressed how temporal constraints and spatial presentation impact processing. I did this by exploring whether the time course and spatial presentation of stimuli impacted how much participants relied on embodied or linguistic factors. First, the time course of an experimental trial was constrained. Results supported integrated theories of cognition with participants relying more on a linguistic factor when given strict time constraints, but when given more time to respond, both linguistic and perceptual factors explained RTs. Furthermore, when participants were told to focus on responding quickly, linguistic factors explained RTs, suggest that task instructions also impact mental representations. In a second experiment, the spatial presentation of stimuli on the screen was modified, and for animate words, decisions were based on the relationship between one word relative to the other words in the experiment. This is in contrast to an explanation where word meaning is based on the relationship between one word and the embodied physical and spatial properties of that simulated word. Together, these experiments implied that both time and space might influence how reliant we are upon both linguistic and perceptual representations during processing. Relying on these findings, I sought to tease apart these two kinds of processing, which are seen as working in parallel. In a third experiment, I demonstrated that both linguistic and perceptual representations, although intertwined are indeed independent processes with linguistic factors predicting performance for semantically related pairs, and only perceptual factors predicting RT performance for unrelated word pairs.

In Chapter 4, I was able to demonstrate that not only is the relative spatial location of stimuli important, but that the orientation of the stimuli is also important. I also demonstrated that gender differences impacted how much participants relied upon linguistic versus perceptual representations. Four experiments demonstrated that primary metaphor processing was best explained by both an embodied cognition and a linguistic account. Further, whether words are presented in a vertical or horizontal configurations modulates how much participants rely on linguistic representations. Finally, female participants were found to be more sensitive to statistical linguistic context than male participants, demonstrating that individual differences also impact whether mental representations.

Chapters 5 and 6 examined prior research by both assessing effect sizes to clarify the conditions under which perceptual simulations are most important and by establishing the number of publications in the current literature that might be reporting incorrect results simply from using an F_1 and F_2 analysis. Chapter 5 determined that effect sizes of language statistics were as large or larger than those of perceptual stimulation with factors associated with immediate processing (button press, word processing) reducing the effect size of perceptual simulation. Chapter 6 established the conditions under which results that are obviously significant for a linear mixed model beget insignificant results for F_1 and F_2 analyses. The results suggested that approximately 34% of the studies using F_1 and F_2 analyses might be subject to a Type I error, with an unknown number of unpublished studies being subject to a Type II error.

The implications of these results are fully in the line with the Symbol Interdependency Hypothesis (Louwerse, 2007) and are highly relevant for the cognitive sciences, as meaning is fundamental to cognitive psychology, linguistics, and computer science, among other fields. In line with theories supporting representational pluralism (Barsalou, et al., 2008; Dove, 2009; Louwerse & Jeuniaux, 2010; Paivio, 1986; Zwaan, 2014) in which conceptual processing blends linguistic and embodied representations, these results similarly suggest that language processing and language grounding is neither exclusively embodied nor exclusively linguistic. Instead mental representations can be both embodied and linguistic, but under different conditions.

The experiments presented in this dissertation explore the intersections of symbolic and embodied theories of language comprehension by examining various factors that determine the extent to which a reader employs symbolic and/or embodied processes during comprehension. In particular, these theoretical perspectives are not presented in contrast against one another, instead, the motivation for these studies was to pursue a unified account of the usefulness of both symbolic and embodied processes.

Adding to the ample evidence that language processing involves the activation of non-linguistic representations, it is also clear from this dissertation that embodiment is not necessarily ubiquitous. Emerging from these results, linguistic predictor variables can explain the very same data that is cited as evidence for embodied cognition. As explicitly stated in Chapter 1, this dissertation does not address an exhaustive set of factors or paradigms that impact mental representations, the goal instead was to provide a clear demonstration that linguistic and embodied factors both play a role in language comprehension. Therefore, it is reasonable to ask whether linguistic frequency and perceptual ratings are an accurate operationalization of symbolic and embodied representations? Although I demonstrated that perceptual ratings and word frequency can be used to estimate the relevance of both symbolic and embodied relations, it is not unreasonable to urge future researchers to expand upon the linguistic and perceptual factors specified here. For example, the experiments presented here focus on embodied representations in the visual modality, it might well be the case that embodied cues from other modalities are more or less relevant to varying extents. Another question that is touched on in this dissertation is what other manipulations might influence language processing. Perhaps additional modulators, tasks, task goals, individual differences, semantic associations, knowledge constraints, expertise, age, etc., impact processing in ways that are yet unknown. For instance, it might be more likely that children would rely less on linguistic cues, as children may lack the necessary language expertise to more efficiently utilize linguistic information than perceptual cues. Although this dissertation asks *when* linguistic and perceptual representations are *more* or *less* relevant during language processing the variables identified here are by no means exhaustive. There are a wide variety of factors that may explain RTs for embodiment experiments, as there are a wide variety of factors that may be manipulated in these experiments.

In line with the symbol interdependency hypothesis (Louwerse, 2007) and the LASS theory (Barsalou, et al., 2008), Chapter 3 also finds that symbolic cognition controls the early stages of comprehension to create a superficial level of representation. Whereas each word could potentially be grounded in the perceptual world, it is not necessary to do so when semantic information can be used to make approximations of meaning. These semantic approximations are able to generate good-enough mental representations. Although linguistic representations are used as early and shallow representations, embodied representations are then employed to process the information more deeply and therefore take longer to generate. For precise mental representations, perceptual simulations of words are necessary. This account implies that linguistic information is important during temporally early processing, and perceptual simulation systems are used later. These findings confirm and replicate the results from Louwerse and Connell (2011) as well of that of Louwerse and Hutchinson (2012). They further extend simple RT findings to hint at task and time constraints that influence when symbolic versus embodied representations are more or less likely to be employed to generate word meaning.

The experiments presented here examine the variables and conditions that impact language processing (with the exception of Chapter 6) and are mostly empirical questions instead of theoretical questions, but they cover a variety of factors (the time course of processing, the spatial presentation of stimuli, individual differences in gender, the orientation of stimuli, task constraints for the participants, differences in stimuli category and type), and show that processing can be shifted towards one type of mental representation or the other, suggesting that under particular conditions, each account alone might have difficult explaining how meaning is assigned to a word.

In conclusion, it is less relevant to consider whether conceptual processing is symbolic or embodied. Instead it is important to determine when, why, and to what extent linguistic and perceptual representations are employed during language processing and under what conditions it is likely that participants will rely more on one type of representation than another. This unified explanation not only predicts a bias towards linguistic factors for linguistic tasks, perceptual factors in perceptual tasks, a bias towards linguistic factors for linguistic stimuli and perceptual factors for perceptual stimuli as previously shown and replicated in Chapter 2, but it also predicts a bias towards linguistic factors in quick tasks, perceptual factors in slow tasks (Chapter 3), linguistic factors in speeded tasks, perceptual factors in accuracy tasks (Chapter 3), linguistic factors in vertical presentation orientations, perceptual factors in horizontal presentation orientations (Chapter 4), linguistic factors for female participants, perceptual factors for male participants (Chapter 4), linguistic factors for related stimuli, and finally perceptual factors for unrelated stimuli (Chapter 3). In sum, a more elegant theory incorporates both linguistic and perceptual factors as opposite sides of the same continuum.

References

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51,* 355-365.

Andreano, J. M., & Cahill, L. (2009). Sex influences on the neurobiology of learning and memory. *Learning and Memory, 16,* 248–266.

Andres, M., Ostry, D., Nicol, F., & Paus, T. (2008). Time course of number magnitude interference during grasping. *Cortex, 44*, 414–419.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological review, 116*, 463-498.

Baayen, R. H. (2008a). *Analyzing linguistic data: A practical introduction to statistics.* Cambridge: Cambridge University Press.

Baayen, R. H. (2008b). LanguageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". R package version 0.953.

Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*. 390–412.

Baayen, R. H. (2001). *Word frequency distributions.* Dordrecht: Kluwer.

Bächtold, D., Baumüller, M., and Brugger, P. (1998). Stimulus-response compatibility in representational space. *Neuropsychologica, 36,* 731-735.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences,* 22, 577–660.

Barsalou, L.W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London: Biological Sciences, 358*, 1177-1187.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617-645.

Barsalou, L. W. (2010). Grounded cognition: Past, present, and future. *Topics in Cognitive Science, 2,* 716–724.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. de Vega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 245–283). Oxford, UK: Oxford University Press.

Barth, H., Kanwisher, N., and Spelke, E. (2003). The construction of large number representations in adults. *Cognition, 86,* 201-221.

Bates, D. M., & Sarkar, D. (2007). Lme4: Linear mixed-effects models using S4 classes, R package version 0.99875-6.

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science, 21,* 1605–1610.

Beilock, S. L., & Holt, L. E. (2007). Embodied preference judgments: Can likeability be driven by the motor system? *Psychological Science, 18(1)*, 51–57.

Benbow, C. P. & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science, 222,* 1029–1031.

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science, 31,* 733–764.

Bergen, B. K., Wheeler, K. B. (2005). Sentence Understanding Engages Motor Processes. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (pp. 1709-1714). Austin, TX: Cognitive Science Society.

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science, 31,* 733-764.

Borghi, A. M., Glenberg, A. M., & Kaschak, M. P. (2004). Putting words in perspective. *Memory & Cognition, 32,* 863– 873.

Bornstein, M. H., Haynes, M. O., Painter, K. M., & Genevro, J. L. (2000). Child language with mother and with stranger at home and in the laboratory: A methodological study. *Journal of Child Language, 27,* 407–420.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition, 75*, 1–28.

Boroditsky, L. (2001). The roles of body and mind in abstract thought. *Psychological Science, 13,* 185–188.

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science, 13,* 185–189.

Borreggine, K. L., & Kaschak, M. P. (2006). The action-sentence compatibility effect: It's all in the timing. *Cognitive Science, 30,* 1097–1112.

Bourke, L., & Adams, A. (2011). Is it differences in language skills and working memory that account for girls being better at writing than boys? *Journal of Writing Research, 3,* 249–277.

Brants, T., & Franz, A. (2006). *Web 1T 5-gram version 1*. Philadelphia: Linguistic Data Consortium.

Brysbaert, M. (2007). *"The language-as-fixed-effect fallacy": Some simple SPSS solutions to a complex problem (version 2.0)*. Royal Holloway, University of London. Technical report.

Burman, D. D., Bitan, T., & Booth, J. R. (2008). Sex differences in neural processing of language among children. *Neuropsychologia, 46,* 1349–1362.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research, 24*, 355–363.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (2001). Spatial-mechanical reasoning skills versus mathematics self-confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items. *Journal for Research in Mathematics Education, 32*, 28–57.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage, 12,* 478-484.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row.

Clark, H. H. (1973). The language-as-fixed effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335–359.

Cohen, J. (1994). The earth is round (*p* < .05). *American Psychologist*, *49*, 997-1003.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Coleman, E. B. (1964). Generalizing to a language population. *Psychological Reports, 14*, 219-226.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A*, 497-505.

Connell, L., & Lynott, D. (2009). Is a bear white in the woods? Parallel representation of implied object color during language comprehension. *Psychonomic bulletin & review, 16*, 573-577.

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin and Review, 20,* 542–550.

Connell, L., Lynott, D., & Dreyer, F. (2012). A functional role for modality-specific perceptual systems in conceptual representations. *PLoS One, 7,* e33321.

Cooper, W. E., & Ross, J. R. (1975). World order. In Robin E. Grossman, L. James San, & Timothy J. Vance (eds.), *Papers from the parasession on functionalism*, (pp. 63–111). Chicago: Chicago Linguistic Society.

Coslett, H. B. (1999). Spatial influences on motor and language function. *Neuropsychologia, 37,* 695-706.

Crawford, L. E., Margolies, S. M., Drake, J. T., & Murphy, M. E. (2006). Affect biases memory of location: Evidence for the spatial representation of affect. *Cognition and Emotion, 20,* 1153–1169.

Davies, C. (2013). Reading geography between the lines: Extracting local place knowledge from text. In T. Tenbrink, J. Stell, A. Galton, & Z. Wood (Eds.), *Conference on spatial information theory (COSIT)* (pp. 320–337). Scarborough, UK: Springer.

Damian, M. F. (2004). Asymmetries in the processing of Arabic digits and number words. *Memory & Cognition, 32,* 164-171.

de Vega, M., Glenberg, A. M., & Graesser, A. C. (eds.). (2008). *Symbols and embodiment: Debates on meaning and cognition*. Oxford, UK: Oxford University Press.

Dehaene, S., Bossini, S., & Giraux, P. (1993) The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122,* 371-396.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition, 43,* 1–29.

Dijkstra, K., Yaxley, R. H., Madden, C. J., & Zwaan, R. A. (2004). The role of age and perceptual symbols in language comprehension. *Psychology and Aging, 19,* 352–356.

Dils, A. T., & Boroditsky, L. (2010). Visual motion after effect from understanding motion language. *Proceedings of the National Academy of Sciences, 107,* 16396–16400.

Dove, G. O. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition, 110,* 412-431.

Estes, Z., Verges, M., & Barsalou, L. W. (2008). Head up, Foot down: Object words orient attention to the objects' typical location. *Psychological Science, 19(2),* 93–97.

Fias, W. (2001). Two routes for the processing of verbal numbers: Evidence from the SNARC effect. *Psychological Research, 65,* 250–259.

Finch, H. (2010). Imputation methods for missing categorical questionnaire data: A comparison of approaches, *Journal of Data Science, 8,* 361-378.

Fischer, M. H., & Brugger, P. (2011). When digits help digits: Spatial– numerical associations point to finger counting as prime example of embodied cognition. *Frontiers in Psychology, 2,* 1–7.

Fischer, M. H., Shaki, S., & Cruise, A. (2009). It takes just one word to quash a SNARC. *Quarterly Journal of Experimental Psychology, 56,* 361–366.

Fodor, J. (1975). *The language of thought*. New York, NY: Crowell.

Fodor, J. A. (2008). *LOT 2: The language of thought revisited.* Oxford, UK: Oxford University Press.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141(1),* 2–18.

Gagné, C. L. (2002). Lexical and relational influences on the processing of novel compounds. *Brain and Language, 81,* 723–735.

Gevers, W., Caessens, B., & Fias, W. (2005). Towards a common processing architecture underlying Simon and SNARC effects. *European Journal of Cognitive Psychology, 17,* 659–673.

Gevers, W., Lammertyn, J., Notebaert, W., Verguts, T., & Fias, W., (2005). Automatic response activation of implicit spatial information: Evidence from the SNARC effect. *Acta Psychologica, 122,* 221-233.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition, 87,* B87– B95.

Gibbs, R. W., Jr. (1994). *The poetics of mind.* Cambridge, UK: Cambridge University Press.

Gibbs, R. W., Jr. (2006). Metaphor interpretation as embodied simulation. *Mind and Language, 21,* 434–458.

Gibbs, R. W., Jr., Lima, P. L. C., & Francozo, E. (2004). Metaphor is grounded in embodied experience. *Journal of Pragmatics, 36,* 1189–1210.

Glenberg, A. M. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and Brain Sciences, 20,* 1–55.

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science, 1,* 586–596.

Glenberg, A. M., Becker, R., Klötzer, S., Kolanko, L., Müller, S., & Rinck, M. (2009). Episodic affordances contribute to language comprehension. *Language and Cognition, 1,* 113– 135.

Glenberg, A. M. & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9,* 558-565.

Glenberg, A. M. & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory & Language*, *43*, pp. 379-401.

Glenberg, A. M., Robertson, D. A., Jansen, J. L., & Johnson- Glenberg, M. C. (1999). Not propositions. *Cognitive Systems Research, 1(1),* 19–33.

Graesser, A.C., McNamara, D.S., Louwerse, M.M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers, 36,* 193-202.

Greenberg, J. H. (1966). *Language universals, with special reference to feature hierarchies.* The Hague: Mouton.

Harnad, S. (1990). The Symbol Grounding Problem. *Physica D*, 42, 335-346.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301–307.

Havas, D. A., Glenberg, A. M., & Rinck, M. (2007). Emotion simulation during language comprehension. *Psychonomic Bulletin & Review, 14,* 436–441.

Holt, L. E., & Beilock, S. L. (2006). Expertise and its embodiment: Examining the impact of sensorimotor skill expertise on the representation of action-related text. *Psychonomic Bulletin & Review, 13,* 694–701.

Howitt, D., & Cramer, D. (2011). *Introduction to research methods*. London: Pearson Education.

Hutchinson, S., Datla, V., & Louwerse, M. M. (2012). Social networks are encoded in language. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 491-496). Austin, TX: Cognitive Science Society.

Hutchinson, S., Datla, V. & Louwerse, M. M. (2012). Social networks are encoded in language. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 491-496). Austin, TX: Cognitive Science Society.

Hutchinson, S., & Louwerse, M. M. (2012). The upbeat of language: Linguistic context and embodiment predict processing of valence words. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1709-1714). Austin, TX: Cognitive Science Society.

Hutchinson, S., & Louwerse, M. (2013). Language statistics and individual differences in processing primary metaphors. *Cognitive Linguistics, 24,* 667–687.

Hutchinson, S., & Louwerse, M. M. (2013b). Language statistics explain the spatial–numerical association of response codes. *Psychonomic Bulletin & Review, 21,* 470–478.

Hutchinson, S., & Louwerse, M. M. (2013c). What's up can be explained by language statistics. In M. Knauff, M. Pauen, N. Sebanz, & I. Washsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 2596–2601). Austin, TX: Cognitive Science Society.

Hutchinson, S., Tillman, R.N., & Louwerse, M.M. (2014). Quick linguistic representations and precise perceptual representations: Language statistics and perceptual simulations under time constraints. *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2399-2404). Austin, TX: The Cognitive Science Society.

IJzerman, H., & Semin, G. R. (2009). The thermometer of social relations: Mapping social proximity on temperature. *Psychological Science, 20,* 1214–1220.

IJzerman, H., & Semin, G. R. (2010). Temperature perceptions as a ground for social proximity. *Journal of Experimental Social Psychology, 46,* 867–873.

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition, 32,* 662–673.

Jostmann, N. B., Lakens, D., & Schubert, T. W. (2009). Weight as an embodiment of importance. *Psychological Science, 20,* 1169–1174.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review, 93,* 1449–1475.

Kan, I. P., Barsalou, L. W., Solomon, K. O., Minor, J. K., and Thompson- Schill, S. L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *Cognitive Neuropsychology, 20*, 525–540.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M. E., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition*, *94*, B79–B89.

Kaschak, M. P., Zwaan, R. A., Aveyard, M., & Yaxley, R. H. (2006). Perception of auditory motion affects language processing. *Cognitive Science, 30,* 733–744.

Kaup, B., Lüdtke, J., & Maienborn, C. (2010). "The drawer is still closed": Simulating past and future actions when processing sentences that describe a state. *Brain and Language, 112,* 159–166.

Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Lüdtke, J. (2007). Experiential simulations of negated text information. *The Quarterly Journal of Experimental Psychology, 60,* 976–990.

Kimura, D. (2000). *Sex and cognition*. Cambridge, MA: The MIT Press.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New York, NY: Cambridge University Press.

Koch, S., Holland, R. W., Hengstler, M., & van Knippenberg, A. (2009). Body locomotion as regulatory process: stepping backward enhances cognitive control. *Psychological Science, 20,* 549–550.

Kövecses, Z. (1986). *Metaphors of anger, pride, and love: A lexical approach to the structure of concepts.* Amsterdam: John Benjamins.

Kövecses, Z. (2005). *Metaphor in culture.* New York, NY: Cambridge University Press. Kuznetsova et al., 2013

Kramer, J. H., Delis, D. C., Kaplan, E., & O'Donnell, L. (1997). Developmental sex differences in verbal learning. *Neuropsychology, 11,* 577–584.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-208.

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2013). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R-Version:1.1-0

Lakens, D. (2011a). High skies and oceans deep: Polarity benefits or mental simulation? *Frontiers in Psychology*, *2,* 1-2.

Lakens, D. (2011b). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bi-polar concepts presented in vertical space. *Journal of Experimental Psychology, 38,* 726-736.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4,* 863.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago, IL: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York, NY: Basic Books.

Landauer, T. K., & Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes, 25,* 259-284.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.

Liederman, J., Kantrowitz, L., & Flannery, K. (2005). Male vulnerability to reading disability is not likely to be a myth. *Journal of Learning Disabilities, 38,* 109–129.

Linn, M. C., & Peterson, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56,* 1479–1498.

Littell, R. C., Stroup, W. W., & Freund, R. J. (2002). *SAS for linear models.* Cary: SAS Institute.

Locker, L., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods, 39(4),* 723–730.

Louwerse, M. M. (2007). Symbolic or embodied representations: A case for symbol interdependency. In T. K. Landauer, D. S. McNamara, S. Dennis & W. Kintsch (eds.). *Handbook of latent semantic analysis*, (pp. 107–120). Mahwah, NJ: Erlbaum.

Louwerse, M. M. (2008). Embodied relations are encoded in language. *Psychonomic Bulletin & Review, 15,* 838–844.

Louwerse, M. M. (2011a). Stormy seas and cloudy skies: conceptual processing is (still) linguistic and perceptual. *Frontiers in Psychology*, *2,* 1-4.

Louwerse, M. M. (2011b). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science, 3,* 273–302.

Louwerse, M. M., & Benesh, N. (2012). Representing spatial structure through maps and language: Lord of the Rings encodes the spatial structure of Middle Earth. *Cognitive Science, 36,* 1556–1569.

Louwerse, M. M., & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science, 35,* 381–398.

Louwerse, M. M., & Hutchinson, S. (2012). Neurological Evidence Linguistic Processes Precede Perceptual Simulation in Conceptual Processing. *Frontiers in Psychology, 3,* 385.

Louwerse, M. M., Hutchinson, S., & Cai, Z. (2012). The Chinese route argument: Predicting the longitude and latitude of cities in China and the Middle East using statistical linguistic frequencies. *Proceedings of the 34th annual conference of the cognitive science society* (pp. 695–700). Austin, TX: Cognitive Science Society.

Louwerse, M. M., & Jeuniaux, P. (2008). How fundamental is embodiment to language comprehension? Constraints of embodied cognition. In V. Sloutsky, B. Love, & K. McRae (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 1313–1318). Austin, TX: Cognitive Science Society.

Louwerse, M. M., & Jeuniaux, P. (2008). Language comprehension is both embodied and symbolic. In M. de Vega, A. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 309–326). Oxford, UK: Oxford University Press.

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition, 114,* 96–104.

Louwerse, M. M., & Zwaan, R. A. (2009). Language encodes geographical information. *Cognitive Science, 33(1),* 51–73.

Maass, A. & Russo, A. (2003). Directional bias in the mental representation of spatial events: nature or culture? *Psychological Science, 14*, 296-301.

Madden, C. J., & Zwaan, R. A. (2006). Perceptual representation as a mechanism of lexical ambiguity resolution: An investigation of span and processing time. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 32,* 1291–1303.

Marques, J.F. (2006). Specialization and semantic organization: Evidence for multiple semantics linked to sensory modalities. *Memory & Cognition, 34,* 60–67.

Markman, A. B., & Brendl, C. M. (2005). Constraining theories of embodied cognition. *Psychological Science, 16(1),* 6–10.

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory and Cognition, 32,* 1389–1400.

McNamara, D., Cai, Z., and Louwerse, M. (2007). Comparing latent and non-latent measures of cohesion. In T. Landauer, D.S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 379-400). Mahwah, NJ: Erlbaum.

Meier, B. P., & Dionne, S. (2009). Downright sexy: Verticality, implicit power, and perceived physical attractiveness. *Social Cognition, 27,* 883–892.

Meier, B. P., Hauser, D. J., Robinson, M. D., Friesen, C. K., & Schjeldahl, K. (2007). What's "up" with god? Vertical space as a representation of the divine. *Journal of Personality and Social Psychology, 93,* 699–710.

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up. *Psychological Science, 15,* 243–247.

Meier, B. P., & Robinson, M. D. (2005). The metaphorical representation of affect. *Metaphor and Symbol, 20,* 239–257.

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion Detection and Motion Verbs: Language Affects Low-Level Visual Perception. *Psychological Science, 18,* 1007–1013.

Meteyard, L., Zokaei, N., Bahrami, B., & Vigliocco, G. (2008). Visual motion interferes with lexical decision on motion words. *Current Biology, 18,* R732–R733.

Myung, J.-Y., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: Manipulation knowledge of objects. *Cognition, 98,* 223–243.

Nayak, N. P., & Gibbs, R. W., Jr. (1990). Conceptual knowledge in the interpretation of idioms. *Journal of Experimental Psychology: General, 119,* 315–330.

Nuerk, H.C., Iversen, W., & Willmes, K. (2004). Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *Quarterly Journal of Experimental Psychology, 57A,* 835–863.

Nuthmann, A., & van der Meer, E. (2005). Time's arrow and pupillary response. *Psychophysiology, 42,* 306–317.

Pacini, A., & Barnard, P. (2011). When the sunny side is down: Re-mapping the relationship between direction and valence. *European Journal of Psychology, 7,* 686–696.

Paivio, A. (1971). *Imagery and verbal processes.* New York, NY: Holt, Rinehart, and Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, UK: Oxford University Press.

Pecher, D., van Dantzig, S., Boot, I., Zanzolie, K., & Huber, D. E. (2010). Congruency between word position and meaning is caused by task Induced spatial attention. *Frontiers in Psychology, 1,* 1–8.

Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *Quarterly Journal of Experimental Psychology*, *62*, 1108-1114.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different modality properties for concepts produces switching costs. *Psychological Science, 14*, 119-124.

Pecher, D., & Zwaan, R. A. (2005). *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*. Cambridge, MA: Cambridge University Press.

Pedhazur, E. (1997). *Multiple regression in behavioral research.* New York, NY: Holt, Rinehart and Winston.

Pinhas, M., & Tzelgov, J. (2012). Expanding on the mental number line: Zero is perceived as the smallest. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1187–1205.

Pinherio, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York, Springer.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 132,* 416–442.

Pylyshyn, Z. (1984). *Computation and cognition: Towards a foundation for cognitive science*. Cambridge, MA: MIT Press.

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59,* 413–425.

Raaijmakers, J. G. W. (2003). A further look at the "language-as-fixed-effect fallacy". *Canadian Journal of Experimental Psychology, 57,* 141–151.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with “the language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language, 41,* 416–426.

Rapp, D. N., & Horton, W. S. (2003). Out of sight, out of mind: Occlusion and the accessibility of information in narrative comprehension. *Psychonomic Bulletin & Review, 10,* 104–110.

Ren, P., Nicholls, M. E. R., Ma, Y., & Chen, L. (2011). Size matters: Non-numerical magnitude affects the spatial coding of response. *PLoS One, 6,* e23553.

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology, 83,* 274–278.

Reynvoet, B., & Brysbaert, M. (1999). Single-digit and two-digit Arabic numerals address the same semantic number line. *Cognition, 72,* 191–201.

Richardson, D., & Matlock, T. (2007). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition, 102(1),* 129–138.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science, 27,* 767-780.

Richter, T., & Zwaan, R. A. (2009). Processing of color words activates color representations. *Cognition, 111,* 383–389.

Roberts, J. K., & Monaco, J. P. (2006). *Effect size measures for the two-level linear multilevel model.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Rueschemeyer S., Glenberg, A. M., Kaschak M. P., Mueller K., Friederici A. D. (2010). Top-down and bottom-up contributions to understanding sentences describing objects in motion. *Frontiers in Psychology, 1,* 1-11.

Rutter, M., Caspi, A., Fergusson, D., Horwood, L. J., Goodman, R., Maughan, B., Moffitt, T. E., Meltzer, H., & Carroll, J. (2004). Sex differences in developmental reading disability. *Journal of the American Medical Association, 291(16),* 2007–2012.

Santana, E., & de Vega, M. (2011). Metaphors are embodied, and so are their literal counterparts. *Frontiers in Psychology,* 2.

Santiago, J., Lupáñez, J., Pérez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review, 14,* 512–516.

Schnall, S., & Clore, G. L. (2004). Emergent meaning in affective space: Conceptual and spatial congruence produces positive evaluations. In Ken Forbus, Dedre Gentner & Terry Regier (eds.). *Proceedings of the twenty-six annual meeting of the cognitive science society*, 1209–1214. Mahwah, NJ: Erlbaum.

Schubert, T. W. (2005). Your Highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology, 89,* 1–21.

Sell, A. J., & Kaschak, M. P. (2010). Processing time shifts affects the execution of motor responses. *Brain and Language, 117,* 39–44.

Semenza, C. (2008). Number Processing. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of the Neuroscience of Language*. (pp. 219-226). London, UK: Academic Press.

Šetić, M., & Domijan, D. (2007). The influence of vertical spatial orientation on property verification. *Language and Cognitive Processes, 22,* 297–312.

Semin, G. R., & Smith, E. R. (Eds.). (2008). *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches*. New York, NY: Cambridge University Press.

Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review, 16,* 328–331.

Shaki, S., & Gevers, W. (2011). Cultural characteristics dissociate magnitude and ordinal information processing. *Journal of Cross-Cultural Psychology, 42,* 639–650.

Simmons, W. K., Hamann, S. B., Harenski, C. N., Hu, X. P., & Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology – Paris, 102,* 106–119.

Spence, C., Nicholls, M. E., & Driver, J. O. N. (2001). The cost of expecting events in the wrong sensory modality. *Attention, Perception, & Psychophysics*, *63*(2), 330-336.

Spivey, M., & Geng, J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research, 65,* 235-241.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12,* 153–156.

Stanovich, K. E. & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23,* 645–665.

Stepper, S., & Strack, F. (1993). Proprioceptive determinants of emotional and nonemotional feelings. *Journal of Personality and Social Psychology*, *64*, 211-220.

Tagalakis, G., & Keane, M. T. (2006). Familiarity and relational preference in the understanding of noun-noun compounds. *Memory & Cognition, 34,* 1285–1297.

Taylor, L., Lev-Ari, S., & Zwaan, R. (2008). Inferences about action engage action systems. *Brain and Language, 107(1),* 62–67.

Tillman, R., Datla, V., Hutchinson, S., & Louwerse, M. M. (2012). From head to toe: Embodiment through statistical linguistic frequencies. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 2434-2439). Austin, TX: Cognitive Science Society.

Tillman, R., Hutchinson, S., Jordan, S. & Louwerse, M. M. (2013). Verifying properties from different emotions produces switching costs: Evidence for coarse-grained language statistics and fine-grained perceptual simulation. In M. Knauff, M. Pauen, N. Sebanz, & I. Washsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3551-3556). Austin, TX: Cognitive Science Society.

Tillman, R., Hutchinson, S., & Louwerse, M. M. (2013). Geographical locations are encoded in statistical linguistic frequencies. In M. Knauff, M. Pauen, N. Sebanz, & I. Washsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3557-3562). Austin, TX: Cognitive Science Society.

Tse, C., Kurby, C. A., & Du, F. (2010). Perceptual simulations and linguistic representations have differential effects on speeded relatedness judgments and recognition memory. *Quarterly Journal of Experimental Psychology, 63*, 928-941.

Tulving, E. (1983). *Elements of episodic memory.* New York, NY: Oxford University Press.

Tulving , E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352-373.

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 166–179.

Van Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2008). Perceptual processing affects conceptual processing. *Cognitive Science, 32,* 579–590.

van Dantzig, S., Zeelenberg, R., & Pecher, D. (2009). Unconstraining theories of embodied cognition. *Journal of Experimental Social Psychology, 45,* 345–351.

Vermeulen, N., Mermillod, M., Godefroid, J., & Corneille, O. (2009). Unintended embodiment of concepts into percepts: Sensory activation boosts attention for same-modality concepts in the attentional blink paradigm. *Cognition, 112,* 467–472.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*, 779–804.

Wagner, S., & Werner, H. (1957). The effect of success and failure on space localization. *Journal of Personality*, *25*, 752-756.

Wei, W., Lu, H., Zhao, H., Chen, C., Dong, Q., & Zhou, X. (2012). Gender differences in children’s arithmetic performance are accounted for by gender differences in language abilities. *Psychological Science, 23,* 320–330.

West B. T., Welch, K. B., Gałecki, A. T. (2006) *Linear mixed models: A practical guide using statistical software.* Boca Raton: Chapman and Hall/CRC.

Wickens, T. D., & Keppel, G. (1983). On the choice of design and of test statistic in the analysis of experiments with sampled materials. *Journal of Verbal Learning and Verbal Behavior*, *22*, 296-309.

Wilson, N. L., & Gibbs, R. W., Jr. (2007). Real and imagined body movement primes metaphor comprehension. *Cognitive Science, 31,* 721–731.

Wilson, A. D. & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology 4,* 58.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499.

Wu, L. & Barsalou, L.W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica, 132,* 173–189.

Yang, F., Mo, L., Louwerse, M. M. (2012). Effects of local and global context on processing sentences with subject and object relative clauses. *Journal of Psycholinguistic Research, 42,* 227–237.

Yaxley, R. H. & Zwaan, R. A. (2006). Simulating visibility during language comprehension. *Cognition, 105,* 229–236.

Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing: The SNARC effect and the REVERSE SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. *Journal of Cognition and Culture, 5,* 165–190.

Zorzi, M., Priftis, K., & Umilta, C. (2002). Brain damage: neglect disrupts the mental number line. *Nature, 417,* 138-139.

Zorzi, M., Priftis, K., Meneghello, F., Marenzi, R., & Umilta, C. (2006). The spatial representation of numerical and non-numerical sequences: Evidence from neglect. *Neuropsychologia*, *44*, 1061-1067.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of Learning and Motivation, 44,* 35–62.

Zwaan, R. A. (2014). Embodiment and Language Comprehension: Reframing the Discussion. *Trends in Cognitive Sciences, 18,* 229–234.

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science, 28,* 611–619.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Do language comprehenders routinely represent the shapes of objects? *Psychological Science, 13*, 168–171.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General; Journal of Experimental Psychology: General*, *135*, 1–11.

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review, 10*, 954–958.

Summary

The cognitive science literature increasingly demonstrates that perceptual representations are activated during conceptual processing. Such findings suggest that the debate on whether conceptual processing is predominantly symbolic or perceptual has been resolved. However, the experiments presented in this dissertation explore the intersections of symbolic and embodied theories of language comprehension by examining various factors that determine the extent to which readers employ symbolic and/or embodied processes during comprehension. In particular, these theoretical perspectives are not presented in contrast to one another, instead, the motivation for these studies was to pursue a unified account of the usefulness of both symbolic and embodied processes. Instead of asking *if* processing relies upon symbolic or embodied representations, the question is posed *when* linguistic and perceptual representations are more or less relevant during language processing, and under what conditions it is likely that participants will rely more on one type of representation than another. More specifically, the question will be addressed to what extent linguistic and perceptual representations are impacted by 1) the time course of processing 2) the spatial presentation of stimuli 3) individual differences or 4) the orientation of stimuli.

Chapter 2 demonstrated that experimental results can be explained by both linguistic and embodied factors. The spatial–numerical association of response codes (SNARC) has shown that parity judgments with participants'

left hands yield faster response times (RTs) for smaller numbers than for larger numbers, with the opposite result for right-hand responses. In three experiments, I replicated the SNARC effect. This effect is traditionally explained in terms of embodied cognition with participants perceptually simulating number magnitude on a mental number line with numbers arranged from small to large. However, this is not the only explanation; in three RT experiments, I showed that the SNARC effect could also be explained by language statistics. Participants made parity judgments of number words (Exp. 1) and Arabic numerals (Exp. 2). Linguistic frequencies of the number words and numbers mirrored the SNARC effect, explaining aspects of processing that a perceptual simulation account could not. Experiment 3 investigated whether high- and low-frequency nonnumeric words would also elicit a SNARC-like effect. Again, RTs were faster for high-frequency words for left-hand responses, with the opposite result for right-hand responses. These results demonstrated that those effects explained solely in terms of perceptual simulation can also be explained by language statistics.

In **Chapter 3** this finding was extended by exploring the time, space, and independence in three experiments. Experiment 1 investigated how the use of linguistic and perceptual representations was impacted when the time course of an experimental trial was constrained. Under time constraints, linguistic frequencies best accounted for participant RTs, but both linguistic and perceptual explanations accounted for slower RTs. Experiment 2 explored how the spatial presentation of stimuli on the screen might also impact how and when participants are more or less likely to rely on linguistic versus perceptual representations. In a RT experiment participants viewed physical-location words at various locations on the screen. For words presented at the top or bottom of the screen, word meaning influenced RTs. But for words appearing in the center of the screen, word frequency played a more important role. In other words, judgments about words were made relative to other words on the screen and not relative to their absolute location on the screen. In a third Experiment I demonstrated that both linguistic and perceptual representations, although intertwined, are relied upon to differing extents based on the nature of the relationship shared between word pairs. In a single RT experiment where participants determined whether linguistically and/or perceptually similar or dissimilar word pairs were semantically related, linguistically related pairs were processed faster than pairs lacking a linguistic relationship whereas perceptually related and unrelated word pairs took longer to process, implying perceptual representation. Furthermore, word frequency predicted RTs for semantically related pairs, whereas both word frequency and perceptual factors were necessary to predict performance for perceptually related pairs. Importantly, for unrelated word pairs, perceptual factors alone predicted RT performance, suggesting that a full perceptual representation is independently utilized when generating a relationship for unrelated word pairs.

Research in cognitive linguistics has emphasized the role of embodiment in metaphor comprehension, with experimental research showing activation of perceptual simulations when processing metaphors. In **Chapter 4**, I discussed how the degree to which linguistic and perceptual information contribute to mental representations varies based on the orientation of the stimuli and on individual differences. In four experiments I showed that language statistics explain the processing of primary metaphors that share an embodied vertical relationship (e.g., *X* above *Y* or *Y* above *X*). Participants saw word pairs with valence, authority, temperature, or gender connotations. The pairs were presented in either a vertical configuration (*X* above *Y* or *Y* above *X*) matching the primary metaphors (e.g., HAPPY IS UP, CONTROL IS UP) or a horizontal configuration (*X* left of *Y* or *Y* left of *X*) not matching the primary metaphors. Results demonstrated that statistical linguistic frequencies explain the response times of the stimulus pairs both in vertical and horizontal configurations, because language has encoded embodied relations. In addition, the effect of the statistical linguistic frequencies was modified by participant gender, with female participants being more sensitive to statistical linguistic context than male participants.

Finally, **Chapter 5** examined effect sizes computed from 126 experiments in 51 previously published embodied cognition studies to clarify the conditions under which perceptual simulations are most important. The effects of language statistics tended to be as large or larger than those of perceptual stimulation and factors associated with immediate processing (button press, word processing) reduced the effect size of perceptual simulation. In **Chapter 6**, I presented a brief discussion with several mathematical simulations to justify my methodological analyses by arguing that linear mixed models provide the most suitable analytical approach to provide answers to the questions posed in this manuscript. I focused on presenting several statistical simulations and explored conditions under which results that are obviously significant for a linear mixed model beget insignificant results for F_1 and F_2 analyses, and vice versa. The second aim of Chapter 6 was to estimate the number of publications in the current literature that might be reporting incorrect results simply from using an F_1 and F_2 analysis.

These chapters demonstrate that research on mental representations can benefit from an integrated viewpoint. I concluded by suggesting that it is less relevant for the cognitive sciences to consider whether conceptual processing is symbolic or embodied and it is instead important to determine when, why, and to what extent linguistic and perceptual representations are employed during language processing.

Journal publications

Hutchinson, S., & Louwerse, M. M. (2015). *Language encodes social network information.* Manuscript submitted for publication.

Hutchinson, S., & Louwerse, M. M. (2015). *Publish or Perish: Consequences of Considering Sampling Errors.* Manuscript submitted for publication.

Louwerse, M. M., Hutchinson, S., Tillman, R., & Recchia, G. (2014). Effect size matters: The role of language statistics and perceptual simulation in conceptual processing. *Language, Cognition and Neuroscience.* 1-18.

Hutchinson, S., & Louwerse, M. M. (2014). Language statistics explains spatial-numerical association of response codes. *Psychonomic Bulletin & Review, 21,* 470-478.

Hutchinson, S., & Louwerse, M. M. (2013). Language statistics and individual differences in processing primary metaphors. *Cognitive Linguistics, 24,* 667 – 687.

Louwerse, M. M., & Hutchinson, S. (2012). Neurological evidence linguistic processes precede perceptual simulation in conceptual processing. *Frontiers in Psychology, 3*: 385.

Refereed Conference Proceedings

Louwerse, M. M., Raisig, S., Tillman, R., & Hutchinson, S. (in press). Time after time in words: Chronology through language statistics. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Tillman, R., Hutchinson, S., & Louwerse, M. M. (in press). How Sharp is Occam's Razor? Language Statistics in Cognitive Processing. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Hutchinson, S., Wei, L., & Louwerse, M. M. (2014). Avoiding the language-as-a-fixed-effect fallacy: How to estimate outcomes of linear mixed models. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Hutchinson, S., Tillman, R., & Louwerse, M. M. (2014). Quick linguistic representations and precise perceptual representations: Language statistics and perceptual simulations under time constraints. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2399-2404). Austin, TX: Cognitive Science Society.

Hutchinson, S., & Louwerse, M. M. (2013). What's up can be explained by language statistics. In M. Knauff, M. Pauen, N. Sebanz, & I. Washsmuth

(Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 2596-2601). Austin, TX: Cognitive Science Society.

Tillman, R., Hutchinson, S., Jordan, S. & Louwerse, M. M. (2013). Verifying properties from different emotions produces switching costs: Evidence for coarse-grained language statistics and fine-grained perceptual simulation. In M. Knauff, M. Pauen, N. Sebanz, & I. Washsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3551-3556). Austin, TX: Cognitive Science Society.

Tillman, R., Hutchinson, S., & Louwerse, M. M. (2013). Geographical locations are encoded in statistical linguistic frequencies. In M. Knauff, M. Pauen, N. Sebanz, & I. Washsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3557-3562). Austin, TX: Cognitive Science Society.

Hutchinson, S., Datla, V. & Louwerse, M. M. (2012). Social networks are encoded in language. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 491-496). Austin, TX: Cognitive Science Society.

Hutchinson, S., & Louwerse, M. M. (2012). The upbeat of language: Linguistic context and embodiment predict processing of valence words. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1709-1714). Austin, TX: Cognitive Science Society.

Louwerse, M. M., Hutchinson, S., & Cai, Z. (2012). The Chinese route argument: Language predicts longitude and latitude of locations. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 695-700). Austin, TX: Cognitive Science Society.

Tillman, R., Datla, V., Hutchinson, S., & Louwerse, M. M. (2012). From head to toe: Embodiment through statistical linguistic frequencies. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 2434-2439). Austin, TX: Cognitive Science Society.

Hutchinson, S., Johnson, S., & Louwerse, M. M. (2011). A linguistic remark on SNARC: Language and perceptual processes in Spatial-Numerical Association. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual meeting of the Cognitive Science Society* (pp. 3437-3442). Austin, TX: Cognitive Science Society.

TiCC PhD Series

1. Pashiera Barkhuysen. *Audiovisual Prosody in Interaction.* Promotores: M. G. J. Swerts, E. J. Krahmer. Tilburg, 3 October 2008.
2. Ben Torben-Nielsen. *Dendritic Morphology: Function Shapes Structure.* Promotores: H. J. van den Herik, E. O. Postma. Co-promotor: K. P. Tuyls. Tilburg, 3 December 2008.
3. Hans Stol. *A Framework for Evidence-based Policy Making Using IT.* Promotor: H. J. van den Herik. Tilburg, 21 January 2009.
4. Jeroen Geertzen. *Dialogue Act Recognition and Prediction.* Promotor: H. Bunt. Co-promotor: J. M. B. Terken. Tilburg, 11 February 2009.
5. Sander Canisius. *Structured Prediction for Natural Language Processing*. Promotores: A. P. J. van den Bosch, W. Daelemans. Tilburg, 13 February 2009.
6. Fritz Reul. *New Architectures in Computer Chess.* Promotor: H. J. van den Herik. Co-promotor: J. W. H. M. Uiterwijk. Tilburg, 17 June 2009.
7. Laurens van der Maaten. *Feature Extraction from Visual Data.* Promotores: E. O. Postma, H. J. van den Herik. Co-promotor: A. G. Lange. Tilburg, 23 June 2009 (cum laude).
8. Stephan Raaijmakers. *Multinomial Language Learning.* Promotores: W. Daelemans, A. P. J. van den Bosch. Tilburg, 1 December 2009.
9. Igor Berezhnoy. *Digital Analysis of Paintings.* Promotores: E.O. Postma, H. J. van den Herik. Tilburg, 7 December 2009.
10. Toine Bogers. *Recommender Systems for Social Bookmarking.* Promotor: A. P. J. van den Bosch. Tilburg, 8 December 2009.
11. Sander Bakkes. *Rapid Adaptation of Video Game AI.* Promotor: H. J. van den Herik. Co-promotor: P. Spronck. Tilburg, 3 March 2010.
12. Maria Mos. *Complex Lexical Items*. Promotor: A. P. J. van den Bosch. Co-promotores: A. Vermeer, A. Backus. Tilburg, 12 May 2010 (in collaboration with the Department of Language and Culture Studies).
13. Marieke van Erp. *Accessing Natural History. Discoveries in data cleaning, structuring, and retrieval*. Promotor: A. P. J. van den Bosch. Co-promotor: P. K. Lendvai. Tilburg, 30 June 2010.
14. Edwin Commandeur. *Implicit Causality and Implicit Consequentiality in Language Comprehension.* Promotores: L. G. M. Noordman, W. Vonk. Co-promotor: R. Cozijn. Tilburg, 30 June 2010.
15. Bart Bogaert. *Cloud Content Contention.* Promotores: H. J. van den Herik, E. O. Postma. Tilburg, 30 March 2011.
16. Xiaoyu Mao. *Airport under Control.* Promotores: H. J. van den Herik, E. O. Postma. Co-promotores: N. Roos, A. Salden. Tilburg, 25 May 2011.
17. Olga Petukhova. *Multidimensional Dialogue Modelling.* Promotor: H. Bunt. Tilburg, 1 September 2011.

18. Lisette Mol. *Language in the Hands.* Promotores: E. J. Krahmer, A. A. Maes, M.G.J. Swerts. Tilburg, 7 November 2011 (cum laude).
19. Herman Stehouwer. *Statistical Language Models for Alternative Sequence Selection.* Promotores: A. P. J. van den Bosch, H. J. van den Herik. Co-promotor: M. M. van Zaanen. Tilburg, 7 December 2011.
20. Terry Kakeeto-Aelen. *Relationship Marketing for SMEs in Uganda.* Promotores: J. Chr. van Dalen, H. J. van den Herik. Co-promotor: B. A. Van de Walle. Tilburg, 1 February 2012.
21. Suleman Shahid. *Fun & Face: Exploring non-verbal expressions of emotion during playful interactions.* Promotores: E. J. Krahmer, M. G. J. Swerts. Tilburg, 25 May 2012.
22. Thijs Vis. *Intelligence, Politie en Veiligheidsdienst: Verenigbare Grootheden?*Promotores: T. A. de Roos, H. J. van den Herik, A. C. M. Spapens. Tilburg, 6 June 2012 (in collaboration with the Tilburg School of Law).
23. Nancy Pascall. *Engendering Technology Empowering Women.* Promotores: H. J. van den Herik, M. Diocaretz. Tilburg, 19 November 2012.
24. Agus Gunawan. *Information Access for SMEs in Indonesia*. Promotor: H. J. van den Herik. Co-promotores: M. Wahdan, B. A. Van de Walle. Tilburg, 19 December 2012.
25. Giel van Lankveld. *Quantifying Individual Player Differences.* Promotores: H. J. van den Herik, A. R. Arntz. Co-promotor: P. Spronck. Tilburg, 27 February 2013.
26. Sander Wubben. *Text-to-text Generation Using Monolingual Machine Translation.* Promotores: E. J. Krahmer, A. P. J. van den Bosch, H. Bunt. Tilburg, 5 June 2013.
27. Jeroen Janssens. *Outlier Selection and One-Class Classification.* Promotores: E. O. Postma, H. J. van den Herik. Tilburg, 11 June 2013.
28. Martijn Balsters. *Expression and Perception of Emotions: The Case of Depression, Sadness and Fear.* Promotores: E. J. Krahmer, M. G. J. Swerts, A. J. J. M. Vingerhoets. Tilburg, 25 June 2013.
29. Lisanne van Weelden. *Metaphor in Good Shape.* Promotor: A. A. Maes. Co-promotor: J. Schilperoord. Tilburg, 28 June 2013.
30. Ruud Koolen. *Need I say More? On Overspecification in Definite Reference.* Promotores: E. J. Krahmer, M. G. J. Swerts. Tilburg, 20 September 2013.
31. J. Douglas Mastin. *Exploring Infant Engagement. Language Socialization and Vocabulary. Development: A Study of Rural and Urban Communities in Mozambique.* Promotor: A. A. Maes. Co-promotor: Dr. P. A. Vogt. Tilburg, 11 October 2013.

32. Philip C. Jackson, Jr. *Toward Human-Level Artificial Intelligence – Representation and Computation of Meaning in Natural Language*. Promotores: H. C. Bunt, W.P.M. Daelemans. Tilburg, 22 April 2014.
33. Jorrig Vogels. *Referential choices in language production: The role of accessibility.* Promotores: A. A. Maes, E. J. Krahmer. Tilburg, 23 April 2014.
34. Peter de Kock. *Anticipating criminal behaviour.* Promotores: H. J. van den Herik, J. C. Scholtes. Co-promotor: P. Spronck. Tilburg, 10 September 2014.
35. Constantijn Kaland. *Prosodic marking of semantic contrasts: do speakers adapt to addressees?* Promotores: M. G. J. Swerts, E. J. Krahmer. Tilburg, 1 October 2014.
36. Jasmina Marić. *Web communities, immigration and social capital.* Promotor: H. J. van den Herik. Co-promotores: R. Cozijn, M. Spotti. Tilburg, 18 November 2014.
37. Pauline Meesters. *Intelligent blauw*. Promotores: H. J. van den Herik, T. A. de Roos.
38. Mandy Visser. *Better use your head: How people learn to signal emotions in social contexts.* Promotores: M. G. J. Swerts, E. J. Krahmer. Tilburg, 10 June 2015.
39. Sterling Hutchinson. *How symbolic and embodied representations work in concert.* Promotores: M. M. Louwerse, E. O. Postma. Tilburg, 30 June 2015.